MPP-2015-13

keV Sterile Neutrino Dark Matter from Singlet Scalar Decays: Basic Concepts and Subtle Features

Alexander Merle* and Maximilian Totzauer†

Max-Planck-Institut für Physik (Werner-Heisenberg-Institut),
Föhringer Ring 6, 80805 München, Germany

June 18, 2015

Abstract

We perform a detailed and illustrative study of the production of keV sterile neutrino Dark Matter (DM) by decays of singlet scalars in the early Universe. In the current study we focus on providing a clear and general overview of this production mechanism. For the first time we study all regimes possible on the level of momentum distribution functions, which we obtain by solving a system of Boltzmann equations. These quantities contain the full information about the production process, which allows us to not only track the evolution of the DM generation but to also take into account all bounds related to the spectrum, such as constraints from structure formation or from avoiding too much dark radiation. In particular we show that this simple production mechanism can, depending on the regime, lead to strongly non-thermal DM spectra which may even feature more than one peak in the momentum distribution. These cases could have particularly interesting consequences for cosmological structure formation, as their analysis requires more refined tools than the simplistic estimate using the free-streaming horizon. Here we present the mechanism including all concepts and subtleties involved, for now using the assumption that the effective number of relativistic degrees of freedom is constant during DM production, which is applicable in a significant fraction of the parameter space. This allows us to derive analytical results to back up our detailed numerical computations, thus leading to the most comprehensive picture of keV sterile neutrino DM production by singlet scalar decays that exists up to now.

*email: amerle@mpp.mpg.de
†email: totzauer@mpp.mpg.de

1 Introduction

Despite great advances in our understanding, our Universe holds mysteries that are yet to be resolved. One of the biggest questions is about the identity of the so-called Dark Matter (DM) which – in terms of cosmic energy density – outweighs ordinary matter by a factor of about five [1,2]. While historically (motivated by, e.g., supersymmetry) our best guess for DM was a Weakly Interacting Massive Particle (WIMP) with a mass of a few hundred GeV and roughly weak interaction strength, by now we have unfortunately not seen a clear signal of such a particle. Even worse, attempts for direct detection disfavour big parts of the parameter space that was deemed "natural" [3–6]. While WIMPs cannot be excluded, we are nevertheless at a point where we should seriously think of alternatives [7].

After all, there are several possibilities left for what DM could be. Alternative ideas range from very light scalars such as axions [8] over non-standard fermions in supersymmetry (such as gravitinos [9] or axinos [10]) up to very heavy exotic options like WIMPzillas [11]. In this work, we concentrate on a candidate motivated by neutrino physics, a sterile neutrino ν_s of (typically) a mass of a few keV. While our natural guess would be for the sterile neutrino mass to be much larger, it is in reality not bound and there exist several consistent theoretical frameworks in which sterile neutrinos can be very light [12, 13]. Sterile neutrino DM has originally been proposed by Dodelson and Widrow (DW) [14] who suggested that, although sterile neutrinos would not have interactions strong enough to keep them in thermal equilibrium in the early Universe, they could nevertheless be produced gradually by the thermal plasma due to their admixtures to active neutrinos. Although the DW mechanism was at that time consistent with all bounds [15], by now we know that it is excluded by structure formation: it produces too hot DM [16]. Taking the sterile neutrino mass to larger values is not possible due to the X-ray bound on the DM decay $\nu_s \to \nu + \gamma$, where ν is any active neutrino (see Ref. [17] for a comprehensive discussion and Ref. [18] for probably the most recent collection of limits).[1]

Several other production mechanisms for keV sterile neutrino DM have been proposed. If a primordial lepton asymmetry is present in the early Universe, the active-sterile transitions could be resonantly enhanced, as proposed by Shi and Fuller (SF) [28]. This mechanism

[1] In 2014 two groups have independently reported a *tentative* signal of an X-ray line at 3.5 keV [19,20] which would, if interpreted as originating from sterile neutrino decay, indicate a particle with a mass of 7.1 keV and an active-sterile mixing angle θ of roughly $\sin^2(2\theta) \sim 7 \cdot 10^{-11}$. However, up to this date there is still a discussion on-going in the literature about whether or not this line is in fact a real signal, see Refs. [21–27]. Given that the signal, if it exists, is at the moment still not at a statistically highly significant level, we do not take any side here but instead suggest to wait until more data is collected, to hopefully either strongly confirm or clearly refute a line signal.

produces a relatively cold DM component in addition to the thermal DW part, which leads to a colder spectrum in accordance with all bounds [17, 29, 30] – however, note that also this mechanism is in some tension with data if the 3.5 keV line is taken seriously [31]. In theories with an extended gauge group, the sterile neutrinos could equilibrate and be produced via generic freeze-out, if the resulting overabundance is corrected by a sufficient amount of entropy production [32, 33] – however this scenario, even though not fully excluded, is on quite general grounds threatened by big bang nucleosynthesis [34].

An alternative is non-thermal production of keV sterile neutrinos via particle decays. This possibility is discussed extensively in the literature, with the decaying particle in most cases being a scalar: variants range from inflaton decay [35, 36] over a general scalar singlet that could itself either freeze-out [37, 38] or freeze-in [39–41] to the case where the scalar is electrically charged [42]. More general possibilities exist as well, for example the production from pion decays [43], from Dirac fermions [44], or from light vector bosons [45, 46]. A benefit of this mechanism is that the velocity spectrum of the keV neutrinos produced by decays is quite generally known to have a tendency to be colder than that from other mechanisms [31, 38, 39, 47, 48].

While several cases of decay production are discussed, the treatments available in the literature involve some crude estimates and approximations. Even though a non-thermal DM spectrum could have interesting and unexpected consequences for cosmological structure formation, many references work on the level of rate equations and only *estimate* the spectrum, if at all. Instead, given that the *distribution function* is the most decisive quantity and that it can be computed from first principles with reasonable effort, its determination should be put on more solid grounds. This is partially done for the production of keV neutrinos by a general singlet scalar entering thermal equilibrium [38], however, only in the approximation where the effective number of relativistic degrees of freedom g_* is constant. While this approximation is not usable in all of the parameter space, it is nevertheless a good approximation for large enough production temperatures and, after all, without this assumption it would not be possible to obtain analytical estimates. Yet, ultimately a fully numerical treatment will be necessary.

In this paper, we will start this endeavour by giving a comprehensive discussion of the production of keV sterile neutrino DM via the decays of general singlet scalars, which themselves are produced via a Higgs portal. As shown in Refs. [38–40], depending on the coupling there exist different regimes where the scalar itself could e.g. be a WIMP or a feebly interacting massive particle and decay in or out of thermal equilibrium. We will generalise the previous treatments and derive the full set of approximate formulas for all regimes possible, but again under the assumption of $g_* = \text{const}$. We furthermore perform a fully numerical study of the allowed regions in the parameter space and obtain distribution functions for all relevant parameter points, which in principle allows us to determine all relevant DM properties for a given point in the parameter space. We determine the regions where the correct DM abundance is obtained for a given sterile neutrino mass, thereby taking into account all relevant bounds.

Structure formation properties of DM candidates can be estimated using the so-called free-streaming horizon $\lambda_{\rm FS}$. With this quantity, we show that different estimates are possible which may lead to quite different results. We clearly illustrate that $\lambda_{\rm FS}$, in spite of being a standard estimator, can lead to inconsistent results depending on the approximations used to calculate it.

The current paper serves as an illustration of the general principles and its purpose is *to give a clear overview of the relatively complicated and subtle details involved.* As such, some aspects lie beyond the scope of the current work. For example, *we neglect active-sterile mixing* and thus completely disregard DW production, which in reality cannot be switched off for non-zero active-sterile mixing (and which on the contrary is desired for the related X-ray signal), in favour of analytical results. Our approximation is still good for very small active-sterile mixing, but it is nevertheless too simplified and will be dropped (as well as the assumption $g_* =$const.) in a future purely numerical investigation of all possible cases [49]. Also the full set of implications for cosmological structure formation cannot be obtained using $\lambda_{\rm FS}$, so that a numerical computation of the structure formation properties is needed. While we could in principle add this to the current work, it would lead away from our main point and furthermore prolong the paper even more, so that we have decided to decouple this study, too [50]. Our goal is to ultimately deliver a fully comprehensive study of scalar decay production of keV sterile neutrino DM, in order to set the stage to discriminate it from alternative mechanisms by future data. Current bounds indicate that this may be possible in the not-too-far future [31], so that this endeavour should be pursued in particular if the 3.5 keV line signal survives.

This paper is organised as follows. We start with an illustrative discussion in Sec. 2, which is supposed to give an overview of our considerations and results at a relatively non-technical level. We then introduce the main equations to be solved in Sec. 3. In Sec. 4, we present analytical approximations for all cases where they can be obtained. After a discussion of the aspects relevant for cosmic structure formation, Sec. 5, we finally present the numerical results of our study in Sec. 6, along with a discussion of all bounds and of the validity of our considerations. We conclude in Sec. 7. Throughout the paper we try to keep the discussion on a minimally technical level, however, all technical details relevant for an inclined reader are exposed in Appendices A and B.

2 Illustrative discussion of the setting and overview of the results

This section mainly serves the purpose of describing in simple terms what we have done in our study. Before entering the technical details, we give an overview not only of the setting we are working in but also of some illustrative results. This will hopefully make the paper more accessible and prevent the reader from getting lost in the technical details.

Our basic setting consists of a real singlet scalar field S which must somehow couple to the right-handed neutrino fields N. The most generic coupling doing this job is a Yukawa term $\frac{y}{2} S \overline{N^c} N$ with coupling strength y which, if the scalar develops a non-zero vacuum expectation value (VEV) $\langle S \rangle$, leads to a Majorana mass $m_N = y \langle S \rangle$, where we have assumed only one generation of right-handed neutrinos for simplicity. Thus, the minimal set of ingredients we need in addition to the standard model (SM) consists of exactly these fields. Our minimal Lagrangian is

$$\mathcal{L} = \mathcal{L}_{\rm SM} + \left[i\overline{N} \partial\!\!\!/ N + \frac{1}{2} \left(\partial_\mu S\right) \left(\partial^\mu S\right) - \frac{y}{2} S \overline{N^c} N + \text{h.c.} \right] - V_{\rm scalar} + \mathcal{L}_\nu \,, \tag{1}$$

which is very similar to what had been used previously [38–40]. Here, $\mathcal{L}_\nu$ denotes the part of the Lagragian giving mass to active neutrinos. The details of this mass generation are in fact not very important for our DM production mechanism, as is the number of fermion generations. However, in the current work, we make the additional assumption of *vanishing active-sterile mixing*. In most settings, there will be couplings between active and sterile neutrinos that in reality cannot be switched off. But, since in this paper we want to present analytical results wherever possible, we take this mixing to be exactly zero and in turn have *no DM production from the DW mechanism.*[2] The scalar potential $V_{\rm scalar}$ takes the most general form compatible with an assumed global $\mathbb{Z}_4$-symmetry:[3]

$$V_{\rm scalar} = -\mu_H^2 H^\dagger H - \frac{1}{2} \mu_S^2 S^2 + \lambda_H \left(H^\dagger H\right)^2 + \frac{\lambda_S}{4} S^4 + 2\lambda \left(H^\dagger H\right) S^2 \,. \tag{2}$$

[2] Given that the astrophysical X-ray bounds push the active-sterile mixing down to tiny values [17,18], this approximation is not necessarily bad. However, an additional contribution from the DW production can modify some of the results obtained here, which is why we will drop this assumption in future works and turn to a purely numerical treatment [49, 50].

[3] For possible issues related to the breaking of this symmetry by a non-vanishing VEV $\langle S \rangle$, see [39] and references therein. Note also that giving up the $\mathbb{Z}_4$ symmetry allows for extra terms in the Lagrangian, resulting in more processes that can contribute to equilibrating the scalar. In this paper, we stick to the symmetry in order to obtain a minimal parameter space for our exploratory analysis. Considerations of taking into account terms linear and cubic in S in the potential can be found in [38].

This potential could easily lead to a VEV $\langle S \rangle \approx \text{GeV} - \text{TeV}$, thus motivating a relatively small Yukawa coupling $y \sim 10^{-9}$–10^{-5} in order for the mass of the sterile neutrino to be in the keV-range.

Not only can the above Yukawa coupling lead to a sterile neutrino mass, but it will also be responsible for sterile neutrino DM production. Several processes could be thought of: the leading contributions are the reactions $SS \leftrightarrow hh$ and $S \to NN$, the first of which couples the scalar field S to the thermal plasma and the second of which produces sterile neutrinos from the decay of S. In principle, also processes like $SS \to NN$ would be possible, but the corresponding rate is proportional to y^4 which is negligible for a sufficiently small Yukawa coupling y. We also neglect the inverse reaction $NN \to S$ which is suppressed due to the heavily suppressed phase space originating in the kinematics of any 2-to-1 process.

Let us add that, in general, there will be mixing between the scalar S and the SM Higgs field after electroweak symmetry breaking, which we have completely ignored. However, given that this mixing is proportional to the generally small Higgs portal coupling λ and that it is also suppressed if the singlet scalar is considerably heavier than the Higgs, which is the case that we are investigating here (cf. Sec. 4), taking into account the mixing would not at all change our results. Note that this would change if one used singlet scalar masses very close to or below the Higgs mass [40], which is why this simplifying assumption must be dropped if we are to extend our considerations to lighter singlets. On the other hand, in that limit we would need to drop further assumptions used in this work (in particular the assumption that g_* is constant), so that it makes sense to postpone these considerations to future work [49].

With these assumptions, it is clear that every scalar present in the early Universe will either decay into two sterile neutrinos or undergo pairwise annihilation into pairs of Higgs bosons. Depending on the exact values of the couplings, there are different regimes possible. For example, if the Higgs portal coupling λ is small enough and the initial abundance[4] of the scalars is zero, then the scalar will never enter thermal equilibrium but it will only be produced occasionally from the plasma. In a more modern language, this mechanism would be called *freeze-in* [51, 52], and the corresponding particle would be called a *feebly interacting massive particle* (FIMP). In this regime, the annihilation into Higgs bosons can be neglected since its reaction rate will be suppressed by the square of the (tiny) scalar density. Hence, the frozen-in abundance of scalars will ultimately be translated into a relic abundance of sterile neutrinos with a particle number just twice

[4] We usually use the term *abundance* to denote a particle number density. Depending on the context, the term *relic abundance* will be used for the particle number density or the corresponding energy density after the process discussed is complete.

the one of the scalars at freeze-in (in a co-moving volume) – irrespective of the size of the Yukawa coupling y as long as it is large enough that all scalars have decayed by now. Another example would be the case where the scalar couples to the Higgs strongly enough to be in thermal equilibrium. Then the argument of doubling the number of particles would still be valid if the decay width of the scalar is sufficiently small for the decay to become effective only after freeze-out. However, if the decay proceeds already while the scalar is in equilibrium, further contributions will be present making the whole picture more complicated. We will discuss all cases in detail in Sec. 4, where we present analytical estimates for the momentum distribution functions f and yields Y wherever possible.

We later on turn to a numerical computation of the DM production. To arrive at the final DM relic abundance, we need to solve a system of Boltzmann equations to compute the momentum distribution function $f(p,t)$ of the DM particle. Ultimately this distribution function contains all relevant information: one can e.g. use it to compute the final DM relic abundance, but it also encodes the information about the evolution of the momentum spectrum with time, i.e., how many particles exist per momentum interval at any given temperature of the Universe. This allows to not only track the course of DM generation in detail, but it furthermore gives information about the velocities of the DM particles at the epochs relevant for the formation of structures (i.e., galaxies and galaxy clusters) in space.

To give a snapshot of the results to be presented, we show in Fig. 1 a plot of the lines of correct DM abundance for several different sterile neutrino masses, in a plane of the parameters $\mathcal{C}_{\rm HP}$ and $\mathcal{C}_\Gamma$ which, as we will explain later, are nothing else than an effective Higgs portal coupling and an effective decay width in convenient units. We have augmented the plot by some example evolutions of the DM production and the underlying distribution functions for some specific parameter points marked in the central figure, in order to illustrate what is behind our computations. The purpose of Fig. 1 is to give a graphical illustration of what will be presented in this paper. All the plots displayed will be discussed in great detail in Sec. 6, where also the terminology, colour-codes, and labelling will be carefully outlined so that, while advancing with the paper, the reader will ultimately be enabled to get a full understanding of Fig. 1.

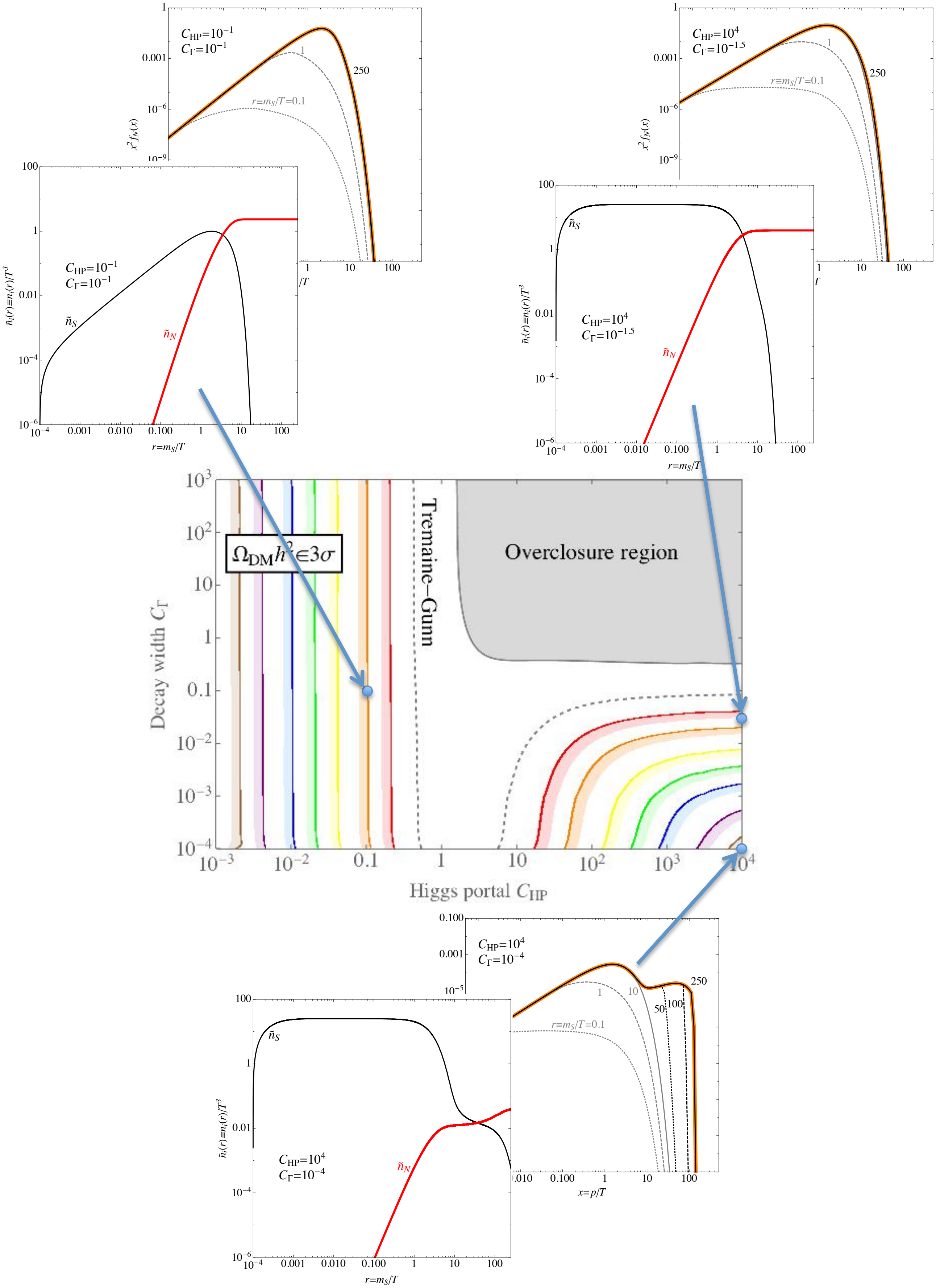


Figure 1: Lines of correct abundance with example distributions and evolutions.

3 The kinetic equations in the early Universe

The fundamental embodiment of every particle species in the early Universe is their distribution function $f(\vec{p},t)$ in momentum space. This quantity does contain all the information we need to deduce the cosmological impact of the species under consideration, so that the determination of f is paramount. Due to isotropy, we will always assume that the distribution functions f only depend on the moduli of the associated 3-momenta. To compute a distribution function, we need to solve the corresponding Boltzmann equation:

$$\hat{L}\left[f\right] = C\left[f\right] . \tag{3}$$

The left-hand side of this equation contains the so-called Liouville operator $\hat{L}$:

$$\hat{L} = \frac{\partial}{\partial t} - Hp\frac{\partial}{\partial p} , \tag{4}$$

where p is the modulus of the particle's 3-momentum and H is the Hubble function. The collision term $C\left[f\right]$ on the right-hand side can be interpreted as a source term, encoding the interaction of the species of interest with itself and with the other particle species present in the plasma. This term contains all the information about the processes which contribute to the production of the species under consideration, and which accordingly shape the resulting momentum distribution function. Collision terms can be relatively lengthy, which is why we report the explicit form of $C\left[f\right]$ only in Appendix A. In this section, however, we prefer to give a more intuitive explanation of how to obtain the distribution functions of the sterile neutrinos.

Since DM production happens in the very early Universe, we only need to consider the era of radiation dominance. During this epoch, the Hubble function can be written as $H\left(T\right) = T^2/M_0$, where T is the temperature of the plasma and M_0 is a function that implicitly depends on time via the evolution of the degrees of freedom g_*:

$$M_0 = \left(\frac{45 M_{\rm Pl}^2}{4\pi^3 g_*}\right)^{1/2} = 7.35 g_*^{-1/2} \times 10^{18}\,{\rm GeV} . \tag{5}$$

Introducing dimensionless variables,

$$x \equiv p/T \qquad \text{and} \qquad r \equiv m_S/T , \tag{6}$$

the Liouville operator from Eq. (4) can be recast into the following form:

$$\hat{L} = \frac{\mathrm{d}r}{\mathrm{d}T}\frac{\mathrm{d}T}{\mathrm{d}t}\frac{\partial}{\partial r} + Hx\left(\frac{\frac{r}{\sqrt{g_*}}\frac{\mathrm{d}}{\mathrm{d}r}\sqrt{g_*}}{1-\frac{r}{\sqrt{g_*}}\frac{\mathrm{d}}{\mathrm{d}r}\sqrt{g_*}}\right)\frac{\partial}{\partial x}\,. \tag{7}$$

Throughout this work, we will stick to the assumption that the numbers of effective degrees of freedom (g_*, g_{*S}) are *constant until the production of sterile neutrinos is completed*, as done e.g. in [38, 39]. This assumption is *absolutely necessary to obtain analytical results* and, in fact, it is not a too bad approximation in a large fraction of the relevant parameter space, as we will illustrate in Sec. 6.5. Nevertheless we will drop it in later works where we are going to present more realistic and purely numerical studies [49, 50].

If the number of effective degrees of freedom does not change during the period of interest, the advantage of the variables x and r becomes obvious. Accordingly, the dynamics of the scalar and the sterile neutrino are given by the following set of equations:

$$\frac{\partial f_S}{\partial r} = \frac{\mathrm{d}T}{\mathrm{d}r}\frac{\mathrm{d}t}{\mathrm{d}T}\left(C^S_{hh\to SS} + C^S_{SS\to hh} + C^S_{S\to NN}\right)\,, \tag{8}$$

$$\frac{\partial f_N}{\partial r} = \frac{\mathrm{d}T}{\mathrm{d}r}\frac{\mathrm{d}t}{\mathrm{d}T}\,C^N_{S\to NN}\,. \tag{9}$$

In Eqs. (8) and (9), the upper indices on the collision terms mark the species the kinetic equation of which is governed by this term, while the subscripts describe the actual process. Thus, $C^S_{SS\to hh}$ describes the depletion of scalars due to annihilations into pairs of Higgses while $C^S_{hh\to SS}$ describes the reverse process. In turn, $C^S_{S\to NN}$ describes the depletion of scalars due to decays and $C^N_{S\to NN}$ encodes the creation of sterile neutrinos from the decays of scalars. Note that the collision terms contain information about the kinematics, too, so that $C^N_{S\to NN}$ and $C^S_{S\to NN}$ differ by more than just a sign. For a detailed derivation and explicit expressions, see Appendix A.

Since we approximate g_* as constant during the time of production and since the matter-radiation equality only takes place at temperatures of $\mathcal{O}\,(1\,\mathrm{eV})$, i.e., long after DM production is completed, we consistently use the time-temperature relation $\frac{\mathrm{d}T}{\mathrm{d}t} = -HT$ in Eq. (9). Using this as well as the explicit form of the collision terms as derived in Appendix A.1, we find the kinetic equation for the scalar:

$$\begin{aligned}\frac{\partial f_S\left(x,r\right)}{\partial r} &= \frac{1}{\sqrt{x^2+r^2}}\left[\frac{1}{4\pi}\mathcal{C}_{\mathrm{HP}}\exp\left(-\sqrt{x^2+r^2}\right)\mathcal{F}\left(x,r\right)-\mathcal{C}_\Gamma r^2 f_S\left(x,r\right)\right.\\ &\qquad\left.-\frac{1}{4\pi}\mathcal{C}_{\mathrm{HP}} f_S\left(x,r\right)\,2\pi\int\limits_0^\infty \mathrm{d}\hat{x}\,\hat{x}^2\int\limits_{-1}^{\alpha_{\max}}\mathrm{d}\cos\theta f_S\left(\hat{x},r\right)\mathcal{G}\left(\hat{x},\cos\theta,r\right)\right]\\ &\equiv \mathcal{Q}\left(x,r\right)-\mathcal{P}\left(x,r\right)f_S-\mathcal{R}\left(x,r\right)\,\mathcal{I}_r\left[f_S\right]\,f_S\,,\end{aligned}\tag{10}$$

where we have defined

$$\mathcal{Q}\left(x,r\right)\equiv\frac{\mathcal{C}_{\mathrm{HP}}\exp\left(-\sqrt{x^2+r^2}\right)\mathcal{F}\left(x,r\right)}{4\pi\sqrt{x^2+r^2}},\tag{11}$$

$$\mathcal{P}\left(x,r\right)\equiv\frac{\mathcal{C}_\Gamma r^2}{\sqrt{x^2+r^2}},\tag{12}$$

$$\mathcal{R}\left(x,r\right)\equiv\frac{\mathcal{C}_{\mathrm{HP}}}{4\pi\sqrt{x^2+r^2}},\tag{13}$$

$$\mathcal{I}_r\left[f_S\right]\equiv 2\pi\int\limits_0^\infty \mathrm{d}\hat{x}\,\hat{x}^2\int\limits_{-1}^{\alpha_{\max}}\mathrm{d}\cos\theta f_S\left(\hat{x},r\right)\mathcal{G}\left(\hat{x},\cos\theta,r\right).\tag{14}$$

For the explicit forms of the kinematic functions $\mathcal{F}$ and $\mathcal{G}$ and the definition of $\alpha_{\max}$, see Appendix A.

In Eqs. (10) to (13), we have used two important dimensionless auxiliary quantities:

$$\begin{cases}\textit{the effective decay width}: & \mathcal{C}_\Gamma\equiv\frac{M_0}{m_S}\frac{\Gamma}{m_S}\,,\\ \textit{the effective (squared) Higgs portal}: & \mathcal{C}_{\mathrm{HP}}\equiv\frac{M_0}{m_S}\frac{\lambda^2}{16\pi^3}\,.\end{cases}\tag{15}$$

In the remainder of this work it will turn out convenient to use these quantities to span the parameter space of our setting. Note that, during the DM production process, we assume M_0 to be constant by virtue of Eq. (5). Hence the interpretations of $\mathcal{C}_{\mathrm{HP}}$ as an effective Higgs portal coupling and of $\mathcal{C}_\Gamma$ as an effective decay width are appropriate, in the sense that the dependence of M_0 on g_* does not change their behaviour and hence they indeed play practically the same roles as the fundamental Lagrangian parameters behind them. For simplicity, we may often refer to $\mathcal{C}_{\mathrm{HP}}$ as *Higgs portal* and to $\mathcal{C}_{\Gamma}$ as *decay width*.

Turning to the sterile neutrino distribution function, which is our main quantity of interest, we can use the assumption of neglecting the back-reaction $NN \to S$ to find a very simple form for the corresponding kinetic equation. Using Eqs. (8) and (9), see Appendix A.2 for details, one can derive a very intuitive *master equation* for the distribution function of the sterile neutrino in terms of that of the scalar:

$$f_N\left(x, r\right) = \int_0^r \mathrm{d}r' \, 2\mathcal{C}_{\Gamma} \frac{r'^2}{x^2} \int_{\hat{x}_{\min}}^{\infty} \mathrm{d}\hat{x} \, \frac{\hat{x}}{\sqrt{\hat{x}^2 + r'^2}} f_S\left(\hat{x}, r'\right), \tag{16}$$

with $\hat{x}_{\min} = \left\| x - r'^2/(4x) \right\|$. It is this equation that will be absolutely central for us: once we have managed to determine the distribution function f_S of the scalar from first principles, we only need to plug it into Eq. (16) to determine the distribution function f_N of the sterile neutrino. As to be expected, the distribution function f_S of the scalar will look differently depending on which regime we are looking at (e.g., the scalar being an early decaying WIMP compared to a late one). This will be translated by Eq. (16) into different resulting sterile neutrino distribution functions. Note that, transforming the momentum variable into an energy, Eq. (16) perfectly coincides with [35, Eq. (8)] and [38, Eq. (10)]. Note also that the master equation for the sterile neutrino distribution is decoupled from the one for the scalar only because of our assumption that the reaction $NN \to S$ is negligible.

4 Analytical results for the distribution functions and relic abundance

In this section, we present the limiting cases that can be treated analytically. Some of the results can be found in the literature, see in particular Ref. [38], while others are completely new. Basically there exist two main cases, the scalar itself can either enter thermal equilibrium in the early Universe and thus act similarly to a WIMP or it can be only very feebly interacting, thus undergoing freeze-in like a FIMP:

1. The *FIMP-regime*: This limiting case is characterised by a small Higgs portal $\mathcal{C}_{\mathrm{HP}}$ and an (almost) arbitrary decay constant $\mathcal{C}_{\Gamma}$. The scalars that are produced via freeze-in subsequently decay into sterile neutrinos. However, for the abundance itself the time of the decay is not very relevant as long as it happens before matter-

radiation equality, which is why the exact value of $\mathcal{C}_\Gamma$ does not matter much in this regime.[5] It will, however, play a role for the spectrum itself, cf. Sec. 6.

2. The *WIMP-regime*: For large enough Higgs portals $\mathcal{C}_{\rm HP}$ the scalar will thermalise, i.e., it enters thermal equilibrium and any information about its initial abundance is lost since the abundance will always be the thermal one. The particle remains in equilibrium until the interaction rate for the process $hh \leftrightarrow SS$ drops below the expansion rate of the Universe and the scalar decouples from the thermal bath. During the course of these events, depending on the exact value of the decay width $\mathcal{C}_\Gamma$, the scalar can decay into sterile neutrinos at various stages:

   (a) *In-equilibrium decay*: If the decay width $\mathcal{C}_\Gamma$ is large enough and if the Higgs portal $\mathcal{C}_{\rm HP}$ is large, too, the sterile neutrinos are produced already very early from the decays of equilibrated scalars. The sterile neutrinos arising from the decays of the few residual scalars that are present after freeze-out can then practically be neglected since the large Higgs portal guarantees a small relic abundance of frozen-out scalars.

   (b) *Out-of-equilibrium decay*: The scalar couples to SM particles strongly enough to enter thermal equilibrium. However, if the decay width is sufficiently small, the production of sterile neutrinos during the time of equilibrium is negligible and it is sufficient to only take into account the late decays of the scalars after their freeze-out. In this regime, the scalar itself acts as a DM-like species, but it is unstable and decays before it can significantly contribute to the energy budget of the Universe.

   (c) *Intermediate regime*: For intermediate values of the decay width $\mathcal{C}_\Gamma$, neither

[5] Note that, for this regime, we will always assume a zero initial abundance for the scalars (and trivially for the sterile neutrinos). If there was a non-negligible initial abundance, this would change our results since the primordial scalars would then add to the ones produced via freeze-in. However, given that there is no reason for such an initial abundance to be present and that we do not see much value in speculating how it could possibly have been produced, we stick to the conservative viewpoint and only produce scalars from the freeze-in mechanism itself.
On the other hand, one could argue that sterile neutrinos and/or singlet scalar fields could quite generically couple to the inflaton field (see Refs. [35,36] for examples concerning the former case). In such scenarios, assuming that inflation is the correct theory in the first place, our scenario might be modified considerably. However, such couplings are only compulsory if the SM gauge group and all other low-energy symmetries are the only ones up to very high scales and even then, from a model building point of view, there exist various possibilities to strongly suppress certain couplings, e.g. by locating the various fields on different branes. While our considerations could in principle be extended to include this point, this would add further complications of which it is however unclear whether they exist, or not. We thus stick to the most minimal setting and disregard any primordial abundance of sterile neutrinos and/or the singlet scalar, as well as the related coupling to the inflaton field.

of the above limiting cases is applicable and it is a priori not clear how well the combination of both of them can correctly describe the situation. On the other hand, this case can be of particular interest since it results into a distribution function with *two* intrinsic momentum scales. This may open up interesting possibilities to tackle the well-known small scale problems of cosmological structure formation [53–56]. We will therefore treat an explicit example for this case numerically in Sec. 5.2.

We will now discuss these different regimes one by one. In all cases, we will take the limit $\xi \equiv m_h/m_S \to 0$, cf. Appendix A.1. Even for $\xi \approx 0.3$, the net effect on the distribution function is of the order of a few per mille. We want to keep the mass m_S of the scalar far from the electroweak (EW) phase transition to ensure that scalars S are only produced by Higgs bosons. For masses $m_S < v_{\rm EW}$, a new range of interaction becomes important like the production of EW gauge bosons from scalars S, $SS \leftrightarrow VV$. This scenario is discussed in [40] on the level of abundances, and we will take these modifications into account on a further more technical study where the focus will be put on even more realistic numerical results [49].

4.1 The FIMP-regime

For sufficiently small Higgs portals $\mathcal{C}_{\rm HP}$ (i.e., for $\lambda \ll 10^{-6}$ [38,39]), the scalar never enters equilibrium and hence one can – for vanishing initial abundance – neglect in Eq. (10) the term $\mathcal{R}\left(r,x\right)\,\mathcal{I}_r\left[f_S\right]\,f_S$ which scales as f_S^2. The remaining kinetic equation is an ordinary differential equation which can be solved analytically. For a vanishing initial abundance of the scalar, the resulting distribution function is given by:

$$f_S\left(r,x\right) = \mathcal{C}_{\rm HP}\int_0^r {\rm d}\rho\;\rho K_1\left(\rho\right)\frac{\exp\left(-\sqrt{\rho^2+x^2}\right)}{\sqrt{\rho^2+x^2}}\left[\frac{e^{\rho\sqrt{\rho^2+x^2}}}{e^{r\sqrt{r^2+x^2}}}\left(\frac{\rho+\sqrt{\rho^2+x^2}}{r+\sqrt{r^2+x^2}}\right)^{-x^2}\right]^{\mathcal{C}_\Gamma/2}, \tag{17}$$

where K_1 is the first modified Bessel function of second kind, cf. Eq. (A-10). Plugging Eq. (17) into Eq. (16) in order to get an analytical result for the distribution function of the sterile neutrino is not very instructive, since there is no simple form of that expression. However, the abundance of the sterile neutrino can be computed analytically for late times, $r \to \infty$. Setting $\mathcal{C}_\Gamma$ to zero corresponds to a stable scalar. With this choice, Eq. (17) can be integrated rather easily and one obtains the (hypothetical) relic abundance of the stable scalar. Since all frozen-in scalars will however ultimately decay into two sterile neutrinos for a non-vanishing decay width $\mathcal{C}_\Gamma$, the abundance of sterile neutrinos will then just be twice the abundance of the would-be-stable scalar [39]. The result for the yield $Y = n/s$, with n and s the particle number and entropy densities, respectively, is given by

$$Y_N\left(r \to \infty\right) = \frac{135}{64\pi^2} \frac{\mathcal{C}_{\mathrm{HP}}}{g_*\left(T_{\mathrm{prod}}\right)} \,. \tag{18}$$

Here, T_{prod} denotes the temperature at the time of production.[6]

4.2 The WIMP-regime

In-equilibrium-decay If both the Higgs portal $\mathcal{C}_{\mathrm{HP}}$ and the decay width $\mathcal{C}_\Gamma$ are large, sterile neutrinos are efficiently produced from the decays of scalars already while being in equilibrium. This case is covered at length in [38]. However, our analytical expression differs by a constant factor, which is why we will sketch the most important steps needed to derive the result. If the scalar is in equilibrium, we know its distribution function exactly. Accordingly, the authors of [38] use a Bose-Einstein (BE) distribution to capture the quantum nature of the scalar. Since the whole set of equations governing the dynamics was however derived using the Maxwellian approximation in the Boltzmann equation, one might wonder whether it would be more consistent to again use a Maxwell-Boltzmann (MB) distribution for the scalar. We will exemplify both cases in order to illustrate that the difference between them is irrelevant.

Plugging the BE and MB distributions into (A-20) yields:

$$f_N\left(x,r\right) = \begin{cases} 8\mathcal{C}_\Gamma \int\limits_1^{z_r} \mathrm{d}z\, x\sqrt{z-1}\log\left(\frac{1}{1-e^{-xz}}\right) & \text{(BE)} \\ 8\mathcal{C}_\Gamma \int\limits_1^{z_r} \mathrm{d}z\, x\sqrt{z-1}e^{-xz} = \frac{e^{-x}\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{x(z_r-1)}\right)}{2\sqrt{x}} - e^{-xz_r}\sqrt{z_r-1} & \text{(MB)} \end{cases} \,, \tag{19}$$

where we have introduced the variable $z_r \equiv r^2/(4x^2)+1$ for convenience.[7] Integrating $f_N\left(x,r\right)$ over d^3x and again taking the limit $r \to \infty$ allows to calculate the yield for late

[6] Of course the time of production is subject to some ambiguities in its definition. Both freeze-in/freeze-out of the scalar and its subsequent decay are continuous processes, the time scales of which are determined by $\mathcal{C}_{\mathrm{HP}}$ and $\mathcal{C}_\Gamma$. It is hence convenient to define the production time as the point when the abundance of sterile neutrinos has passed some threshold fraction of the final abundance, which we take to be 95%.

[7] Note that, as we had already mentioned, we have neglected the mixing between the two physical scalars, which is a very good approximation in our case. However, in order to simplify the comparison of our results to the ones obtained in Ref. [38], it is of course necessary to apply the same approximation to the results from that reference.

times:

$$Y_N\left(r\to\infty\right)=\begin{cases}\frac{135}{4\pi^3}\zeta\left(5\right)\frac{\mathcal{C}_\Gamma}{g_*\left(T_{\rm prod}\right)} & \text{(BE)}\\ \frac{135}{4\pi^3}\frac{\mathcal{C}_\Gamma}{g_*\left(T_{\rm prod}\right)} & \text{(MB)}\end{cases}\,. \tag{20}$$

Both results only differ by a factor of $\zeta\left(5\right)\approx 1.0369$, which justifies the use of either distribution. Our result in the BE case is however larger by a factor of $5/2$ compared to the one reported in [38]. While one may easily forget powers of two in these computations, we could not trace any step where a factor of 5 could possibly be introduced, making us confident that our results are correct.

Out-of-equilibrium decay This limiting case describes a scenario where the scalar is in equilibrium and ultimately freezes out. The decay width $\mathcal{C}_\Gamma$ is so small that practically no sterile neutrinos are produced before the scalar decouples from the plasma. Only after the scalar has frozen-out, the production of the sterile neutrinos sets in. As in [38], we make the approximation that the scalar has a thermal distribution until it freezes out instantaneously at $r=r_{\rm FO}$. In this case the kinetic equation, Eq. (10), can be solved:

$$f_S\left(x,r>r_{\rm FO}\right)=f_{\rm eq}\left(x,r_{\rm FO}\right)\left(\frac{r+\sqrt{r^2+x^2}}{r_{\rm FO}+\sqrt{r_{\rm FO}^2+x^2}}\right)^{\mathcal{C}_\Gamma x^2/2}e^{-\mathcal{C}_\Gamma\left(r\sqrt{x^2+r^2}-r_{\rm FO}\sqrt{x^2+r_{\rm FO}^2}\right)/2}\,, \tag{21}$$

where $f_{\rm eq}\left(x,r_{\rm FO}\right)$ is the equilibrium distribution of S at freeze-out. It could again be taken to be BE or, more consistently, MB. In the case of BE, our expression coincides with [38, Eq. (43)]. The final abundance of sterile neutrinos can in this limiting case again be calculated from doubling the abundance of scalars at freeze-out. Given as a function of $r_{\rm FO}$, the expression for the yield is

$$Y_N\left(r\to\infty\right)=\frac{45}{4\pi^4 g_*\left(T_{\rm prod}\right)}\int\limits_{r_{\rm FO}}^{\infty}{\rm d}\epsilon\,\epsilon\frac{\sqrt{\epsilon^2-r_{\rm FO}^2}}{e^\epsilon-\delta}\quad\left(=\frac{45r_{\rm FO}^2K_2(r_{\rm FO})}{4\pi^4 g_*\left(T_{\rm prod}\right)}\text{ for MB}\right)\,, \tag{22}$$

with $\delta=1$ ($\delta=0$) for the BE (MB) case. Also here the numerical difference between both versions is fairly small for realistic values of $r_{\rm FO}$.

Intermediate regime If $\mathcal{C}_\Gamma$ is in an intermediate regime, neither the production *before* nor the one *after* freeze-out completely dominates the sterile neutrino distribution function, such that no simple analytical treatment is possible. In this instant we have to rely on a purely numerical treatment. We will present a sample case for this regime in

Sec. 6. Nevertheless this intermediate regime may be of special interest as it features *two* intrinsic scales.

4.3 Calculating the relic abundance

So far we have shown formulae to compute the yield at $r \to \infty$. To convert this into the commonly quoted closure parameters, we first have to multiply the yield by today's entropy density of $s_0 = 2891.2\ \text{cm}^{-3}$ [57] and then compare the DM energy density today to the critical density of the Universe. We can write

$$\Omega_{\text{DM}} h^2 = \frac{m_N Y_N(r \to \infty) s_0}{\rho_{\text{crit}}/h^2}\,, \tag{23}$$

where $\rho_{\text{crit}}/h^2 = 1.054 \times 10^{-2}\,\text{MeV cm}^{-3}$ [57]. With this conversion formula at hand, it is straightforward to transform the analytical estimates for the yield Y_N in Secs. 4.1 and 4.2 into expressions for the DM relic abundance, which can then be compared to the observed 1σ range obtained by the Planck collaboration, $\Omega_{\text{DM}} h^2 = 0.1188 \pm 0.0010$ [2].

5 Aspects of structure formation

Up to now, we have only been concerned with the DM relic abundance, i.e., the *amount* of DM in the Universe. However, for a viable DM candidate it is not only important to be sufficiently abundant in the Universe, but it must also allow for structures such as galaxies to form. Clearly, this imposes constraints on the momentum distribution function $f(p,t)$ of the DM candidate under consideration: it must not be hot, i.e., it is not allowed to be too relativistic at the time of structure formation when galaxies form (more precisely, the allowed amount of hot DM is at most a tiny 1% of all DM in the Universe [58, 59]).

The form of the distribution function $f(p,t)$, i.e. the spectrum of the DM candidate, can be constrained from the observed matter distribution in the cosmos: the evolution of spatial inhomogeneities depends on the spectrum of the DM particles and therefore it ultimately constrains the DM production mechanism. Since it is computationally impossible to run a simulation of structure formation for any possible distribution function, a commonly used indicator that can be calculated easily for a given $f(p,t)$ is the so-called *free-streaming horizon* λ_{FS}. This quantity describes the average distance a DM particle would have travelled after production without collisions and not subject to gravitational clustering. In fact this quantity can serve as a good estimator for a length below which the formation of structures is heavily suppressed. For that reason, the free-streaming horizon $\lambda_{\rm FS}$ is commonly used in the literature to discard DM models with a spectrum that is too hot to explain the filamentary structure of the large scale matter distribution in the Universe by preventing galaxy-sized objects of being formed.

5.1 Treatment of the free-streaming horizon

In this section, we will show that the free-streaming horizon itself is subject to substantial uncertainties in its definition, which will make clear why we later present some of our results in two different versions. Moreover we will demonstrate that even our simple one-component DM model can produce highly non-thermal momentum distribution functions, which may even feature *two* intrinsic momentum/velocity scales. In such a case, the average momentum (and hence the free-streaming horizon) cannot be expected to lead to sensible conclusions.

The free-streaming horizon is defined by [16]:

$$\lambda_{\rm FS} \equiv \int_{T_{\rm prod}}^{T_0} \frac{\langle v\left(T\right)\rangle}{a\left(T\right)} \frac{{\rm d}t}{{\rm d}T} {\rm d}T\,, \tag{24}$$

where $T_{\rm prod}$ is the temperature at which production can be seen as complete (in the sense discussed before) and T_0 is today's temperature. Here, $\langle v\left(T\right)\rangle$ is the average velocity of the sterile neutrinos that can be calculated from the distribution function:

$$\langle v\left(T\right)\rangle = \frac{\int\limits_0^\infty {\rm d}x\ \frac{x^3}{\sqrt{x^2+r^2(m_N/m_S)^2}} f_N\left(x,r\right)}{\int_0^\infty {\rm d}x\ x^2 f_N\left(x,r\right)}\,. \tag{25}$$

From Eq. (25) it becomes clear that $\langle v\rangle$ converges to unity, i.e., to the speed of light, as long as f_N is concentrated around values of $x = p/T \gg \sqrt{r^2 m_N^2/m_S^2} = m_N/T$, just as expected. Since $f_N\left(x\right)$ does not change after the production process is complete, the factor of $r^2 m_N^2/m_S^2$ in the square root increases. Note also that, once $r^2 m_N^2/m_S^2$ is greater than the value(s) of x around which f is concentrated, Eq. (25) converges to the non-relativistic expression, $\langle v\rangle \to T/m_N \langle x\rangle \equiv \langle p\rangle/m_N$.

For our numerics, we follow the approximations usually found in the analytical approach [16], namely we will assume the Universe to be completely radiation dominated until some temperature $T_{\rm eq}$ ("matter-radiation equality"), where the Universe switches to being completely matter dominated. The last epoch of vacuum dominance is irrelevant:

even though this period dominates the past of the Universe on an absolute time scale (the matter-vacuum equality being located at a redshift of $z = \Omega_\Lambda/\Omega_{\rm m} \approx 2.2$, corresponding to a time of roughly 3×10^9 a [60]), the velocities of the sterile neutrinos in this epoch are so tiny that the resulting contribution to the free-streaming horizon is negligible. Commonly, the evolution of the degrees of freedom (d.o.f.) is implemented by an additional dilution factor of $\xi^{1/3} = [g_{*S}(T_{\rm prod})/g_{*S}(T_0)]^{1/3} \approx (106.75/3.36)^{1/3} \approx 3.17$, by which $\lambda_{\rm FS}$ is rescaled to account for entropy dilution (cf. [38, 39]). Note that, although we approximate $g_{*S} \approx$ const. during DM production, we nevertheless have to take into account the entropy dilution until today, since there is no justification for the above assumption to be valid through the entire history of the Universe. Departing from the above approximation also during DM production would modify the dilution factor $\xi^{1/3}$, but not too drastically because of the presence of the third root. We will later on perform an estimate of the validity of our approximation, cf. Sec. 6.5. Note that, however, in this formalism the dependence on $g_{*S}\,(T_{\rm prod})$ is quite mild, such that it is safe to use the SM number of d.o.f. $g_{*S} = 106.75$ even though the new particles contribute as well to some degree, depending on how strongly suppressed their true distribution functions are compared to a thermal one.

There is, however, an issue with the analytical approach. The above treatment does not capture the physical fact that, in reality, the evolution of the d.o.f. also enters in the time-temperature relation inside the integral in Eq. (24), but this can be taken into account rather easily in a numerical evaluation of the integral. We therefore compute $\lambda_{\rm FS}$ in a second (numerical) version, in order to take the full evolution of the d.o.f. into account. Thereby, our numerics uses a set of fitted analytical formulae for the evolution of the d.o.f. [61]. For more technical details on this numerical integration, see Appendix B.

In order to compare to the results already present in the literature, we will follow both approaches in parallel. To get an idea about whether the DM can be classified as cold, warm, or hot for a certain set of parameters, we take $\lambda_{\rm FS} \overset{!}{=} 0.1\,{\rm Mpc}$ to mark the boundary between hot and warm DM. This choice is relatively common in the literature (see, e.g., Refs. [62–65]), and it corresponds to the size of a typical dwarf satellite galaxy, which yields a sensible physical motivation. The boundary between warm and cold DM in turn is smooth but it is clear that $\lambda_{\rm FS}$ should be considerably smaller in this case, so that we choose $\lambda_{\rm FS} \overset{!}{=} 0.01\,{\rm Mpc}$, i.e., one order of magnitude smaller than for the hot/warm boundary. Of course there is some arbitrariness involved in these choices, but given that the free-streaming horizon in itself can only yield an indication, the actual error introduced is not as serious as it may seem at first sight. In general the free-streaming horizon can only serve as an order-of-magnitude estimate, and it clearly should not be used to prematurely discard unclear results. Ultimately, only more advanced computations of structure formation can assess whether scenarios with borderline free-streaming horizons should be discarded, or not [50]. For recent developments beyond the commonly used free-streaming approach, see e.g. [66].

5.2 Free-streaming horizon failing – an explicit example featuring twin peaks

As mentioned before, even our simple one-component DM model can feature a distribution function with *two* intrinsic scales, namely in the case where the relic abundance of sterile neutrinos is produced from the decay of equilibrated scalars and the decay of frozen-out scalars in comparable amounts. We present here the exemplary case of $\mathcal{C}_{\rm HP} = 10^4$ and $\mathcal{C}_\Gamma = 10^{-3.5}$, which yields the correct relic abundance for a (relatively large) sterile neutrino mass of about 73 keV. Fixing the scalar mass to be 1 TeV and the number of degrees of freedom at high tempertures to $g_* = 106.5$, this would correspond to values of the Lagrangian couplings of $(y, \lambda) = (4.7 \times 10^{-9}, 8.3 \times 10^{-5})$.

Fig. 2 shows the distribution function of the sterile neutrino for different values of the time parameter r. One can clearly see how early production from the plasma populates the lower comoving momenta (dubbed *Thermal part* – although the resulting distribution may not be perfectly thermal), while the late contributions mainly originate from the decay of the frozen-out scalars (*Decay part*). The inset in the plot shows the enlarged region between the two peaks.

It is obvious that in this case the average momentum does not at all capture the characteristics of the distribution function. According to our estimates using $\lambda_{\rm FS}$ and the average value $\langle x \rangle \equiv \langle p \rangle / T \simeq 16.6$, this point in the parameter space corresponds to a scenario where the sterile neutrinos are on the borderline between being hot and being warm (cf. Sec. 6). However, in fact the low momentum ("thermal") part with $\langle x \rangle_{\rm low} \approx 2.5$ contributes practically as strongly as the high momentum ("decay") part with $\langle x \rangle_{\rm high} \approx 35.7$, where in both cases we have approximated the respective peaks with the individual results from Eqs. (19) and (21), respectively. Note that this splitting introduces some numerical uncertainties, since the expression in Eq. (21) is quite sensitive to small deviations in $r_{\rm FO}$, which in turn suffers from some arbitrariness in the exact definition (due to freeze-out being a process with a small but finite temporal extent rather than an immediate effect).

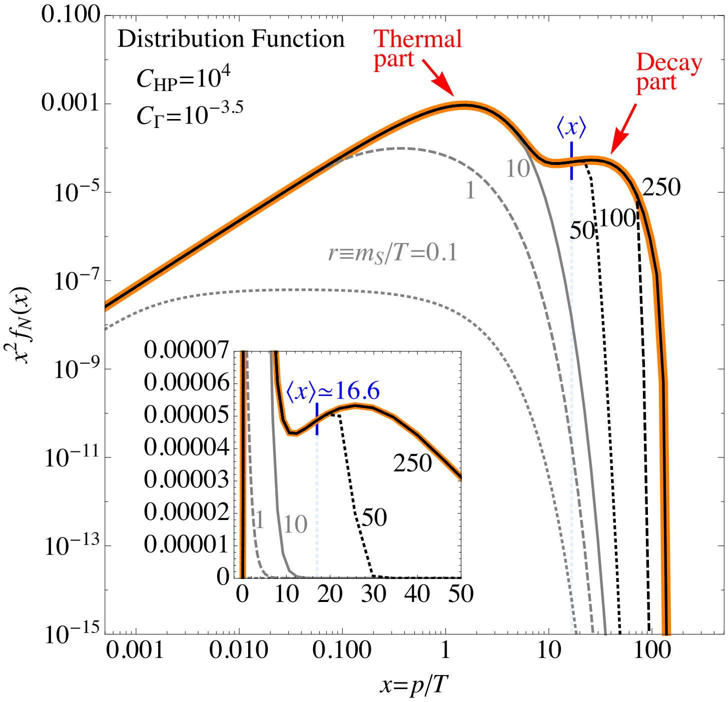


Figure 2: Example of the evolution of a distribution function of sterile neutrinos. One can clearly distinguish two momentum scales (global maximum at $x_1 \approx 1.5$ and second, local maximum at $x_2 \approx 26$). The mean comoving momentum $\langle x \rangle$ is located at $\langle x \rangle \approx 16.6$. The standard deviation $\sqrt{\langle x^2 \rangle - \langle x \rangle^2}$ is approximately 26, which illustrates that the mean value $\langle x \rangle$ contains practically no information.

Estimating the two corresponding free-streaming horizons, cf. Eq. (24),

$$\left(\lambda_{\text{FS,thermal}}^{\text{numerical}}, \lambda_{\text{FS,thermal}}^{\text{estimate}}\right) \sim (0.05, 0.01) \text{ Mpc}\,,$$

$$\left(\lambda_{\text{FS,decay}}^{\text{numerical}}, \lambda_{\text{FS,decay}}^{\text{estimate}}\right) \sim (0.7, 0.1) \text{ Mpc}\,,$$

the left peak tends to yield cold/warm DM while the right one would indicate hot DM in both cases. The full distribution function corresponds to $\left(\lambda_{\text{FS,total}}^{\text{numerical}}, \lambda_{\text{FS,total}}^{\text{estimate}}\right) \approx (0.27, 0.05)$ Mpc, which is perfectly in between the two individual estimates and consistently indicates a case at the borderline of warm/hot DM. Even though the splitting into two distinct parts introduces some extra numerical uncertainty with respect to the values for the complete distribution function, these values clearly illustrate the issue with using the free-streaming horizon as an estimator.

This is precisely one of the cases where more detailed studies about structure formation

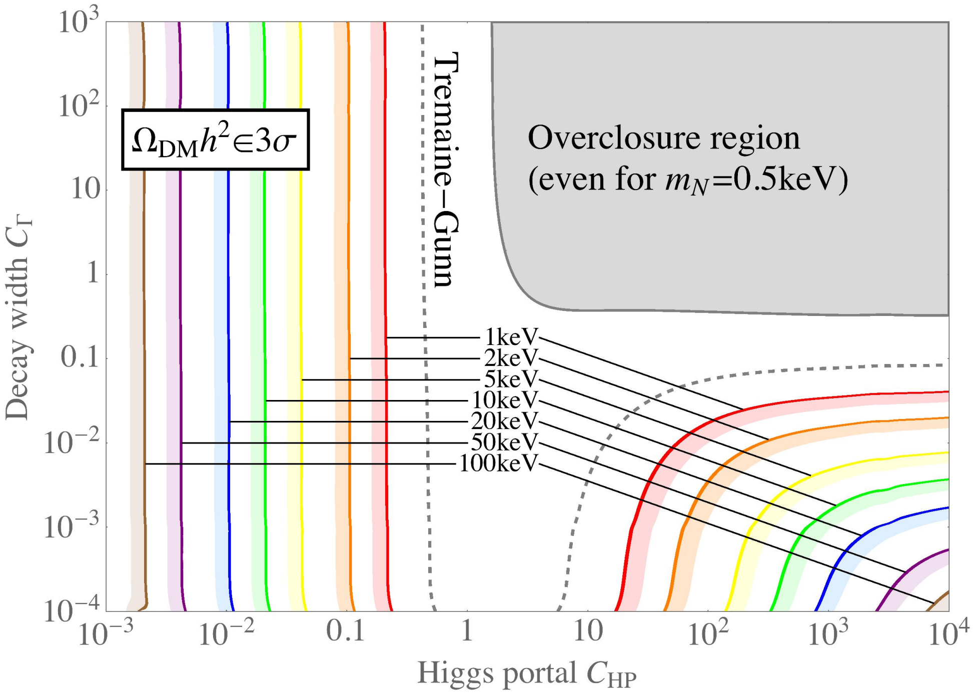


Figure 3: Lines of correct abundance (dark colours) and of sizeable but insufficient abundance (faint colours) in the $\mathcal{C}_{\rm HP}$-$\mathcal{C}_\Gamma$ plane for different values of the sterile neutrino mass. In addition, the Tremaine-Gunn and overclosure bounds are displayed, see text for details.

have to be performed in order to obtain a final answer. With the simple estimate of the free-streaming horizon alone, such a scenario should not be prematurely discarded. In any case, it is worthwhile noting that a one-component DM model can have two similarly important momentum scales in its distribution function, which may open up new possibilities to tackle the small-scale problems from structure formation simulations.

6 Results and bounds

In this section we will present our detailed results and we will also address possible bounds on the scenario as well as the validity of our considerations.

6.1 The Dark Matter abundance

Let us first discuss the DM abundance, which is displayed in Fig. 3 (similarly to the one shown in Sec. 2). In this plot, the dark coloured bands mark the regions in the $\mathcal{C}_{\rm HP}$-$\mathcal{C}_\Gamma$ plane where a sterile neutrino of a given mass yields an abundance in accordance with the 3σ range from the Planck 2015 data [2]. For example, the dark red lines correspond to a sterile neutrino with a mass of $m_N = 1$ keV. Here, the lines on the left correspond to the FIMP regime while the ones on the right correspond to the WIMP regime. We also mark, by the neighbouring fainter lines, the regions where a sizeable but not sufficiently large abundance is generated.

There are two important bounds displayed in Fig. 3. Let us start with the so-called Tremaine-Gunn (TG) bound [67], which is based on the idea that any collection of identical fermions must have a certain minimum phase space density. Applying this bound to the observed dwarf satellite galaxies leads to a lower bound of roughly $m_N > 0.5$ keV on the sterile neutrino mass [68]. Thus, smaller masses are ultimately forbidden by the Pauli exclusion principle. In our plots this bound is marked by the gray dashed line around the white area which in particular cuts into the parameter space for values of $\mathcal{C}_{\rm HP}$ close to one. It basically indicates where the relic density for $m_N = 0.5$ keV equals the upper 3σ bound [2]. The second bound comes from the fact that the DM should not "overclose" the Universe, i.e., its energy density fraction $\Omega_{\rm DM}$ should be smaller than one. Since in the figures we marked the lines of correct abundance for different masses m_N, the resulting forbidden regions do in fact also depend on the mass of the DM particle. However, given that there is a model-independent lower value for the mass from the TG bound, at least for this smallest mass of $m_N = 0.5$ keV the overclosure region marks an absolute bound. Not too surprisingly, this forbidden region is smaller than that excluded by the TG bound, and in our plots it is marked by the gray patches in the upper right corners.

As already indicated in Secs. 2 and 4, depending on the exact values of the parameters different regimes are possible. Let us discuss a few numerical examples. Starting with the FIMP case, in Fig. 4 we illustrate an example point $(\mathcal{C}_{\rm HP}, \mathcal{C}_\Gamma) = (10^{-1}, 10^{-1})$ which corresponds to this regime. Taking again the benchmark value $m_S = 1\,\text{TeV}$ and using $g_* = 106.5$ at high temperatures, the effective couplings translate into $(y, \lambda) = (8.4 \times 10^{-8}, 2.6 \times 10^{-7})$ on the Lagrangian level. On the left panel, we depict the evolution of the sterile neutrino distribution function $f_N(r)$ with the time parameter r. As one can see, most of the abundance is produced around the time $r \sim 1$. Soon afterwards the production ceases such that even for very late times, the distribution hardly changes (as soon as $r \sim 10$, the distribution is practically identical to the final one). The distribution

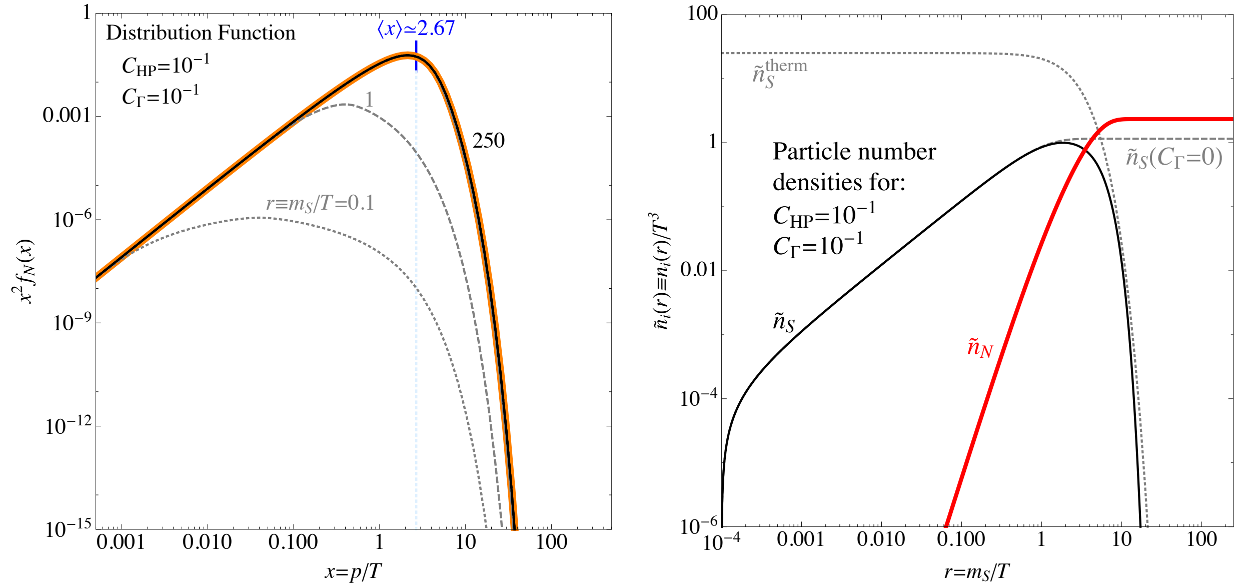


Figure 4: The evolutions of the distribution function (left) and of the abundances (right) for a point corresponding to the scalar being a FIMP.

exhibits a clear peak whose maximal value is very close to the mean momentum over temperature, $\langle x\rangle \simeq \langle p/T\rangle \simeq 2.67$. This means that this sterile neutrino distribution is *colder* than a thermally produced one, for which this number would be equal to 3.15 [35]. However, this point nevertheless turns out to be in the hot DM region, cf. Sec. 6.2. This is mainly due to the fact that this point in parameter space requires a mass of the sterile neutrino of $m_N \approx 2\,\mathrm{keV}$ in order to fulfil the relic abundance constraint. This low mass in turn leads to a long time of highly relativistic free-streaming.

The evolution of the abundances $\tilde{n}_i$, i.e., the integrals over the distribution function divided by T^3, is depicted on the right panel of Fig. 4. We display both the abundances $\tilde{n}_S$ of the scalar and $\tilde{n}_N$ of the sterile neutrino. Starting with the scalar (solid black line) we can see that, as to be expected from a generic FIMP, the abundance of the scalar is gradually built up with increasing time parameter r. However, it never reaches a thermal abundance, as can be seen by comparing the black curve to the hypothetical one for a scalar in thermal equilibrium (dotted gray line). If the scalar was stable, its abundance would not anymore change after the freeze-in is completed, cf. dashed gray line, and it would in practice act as some type of DM. However, given that the scalar is unstable, once it does decay its abundance decreases and instead a sizeable abundance $\tilde{n}_N$ of sterile neutrinos is built up (red solid line). Indeed, because of each scalar decaying into two sterile neutrinos, the final abundance of sterile neutrinos is exactly twice the one of the would-be-stable scalar for late times (numerically we obtain $\tilde{n}_N(r = 250)/\tilde{n}_S(\mathcal{C}_\Gamma = 0, r = 250) \simeq 2.03$, in excellent agreement with the expectation). Furthermore, we can use Eq. (18) to cross-check our numerics, and both results agree nearly perfectly with each other, within a deviation of only 2.8% in this case. Note that this value is *not* a measure of the quality of our numerical methods since we do not know a priori in which part of the parameter space the analytical results approximate the exact result to a desired accuracy. We expect the deviation to become smaller as $\mathcal{C}_\Gamma$ further decreases. In fact, on the edge of our parameter space where $\mathcal{C}_\Gamma = 10^{-4}$, the analytical result is reproduced with deviations well below 1%.

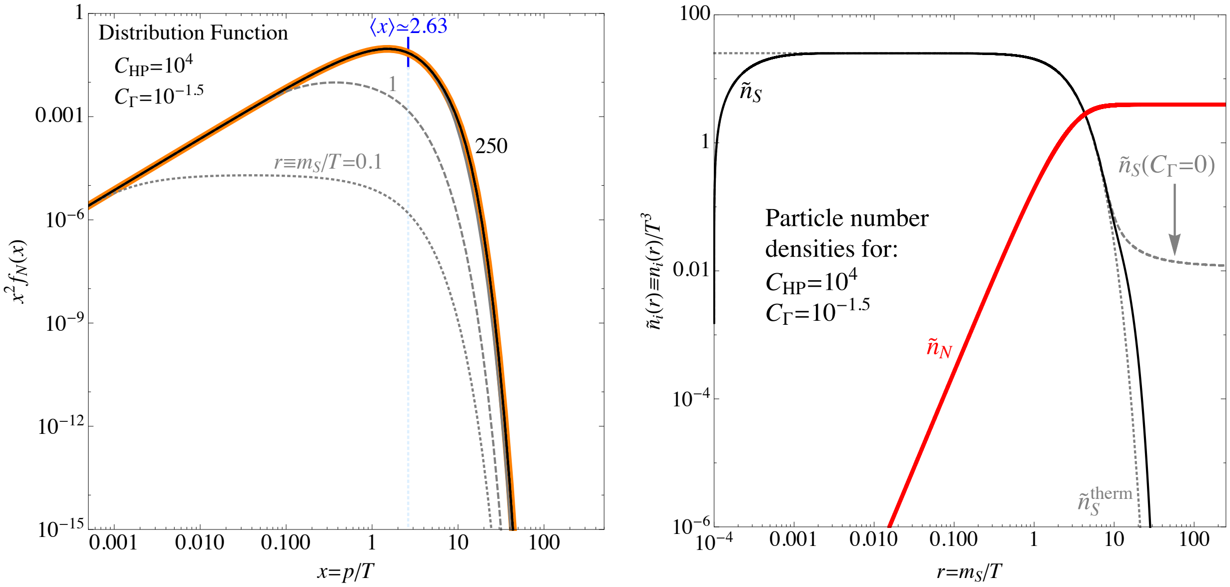


Figure 5: The same as Fig. 4, but for the WIMP case with decay in equilibrium.

Let us now turn to the WIMP cases. For a large enough decay width, an equilibrated scalar can decay while being in thermal equilibrium. This regime is in fact only realised in a relatively small corner of the parameter space, but one point which is a good example for this behaviour is $(\mathcal{C}_{\rm HP}, \mathcal{C}_\Gamma) = (10^4, 10^{-1.5})$ – corresponding to $(y, \lambda) = (4.72 \times 10^{-8},\, 8.3 \times 10^{-5})$ for our standard reference values of $m_S = 1\,\text{TeV}$ and $g_* = 106.5$ – as depicted in Fig. 5. Glancing at the distribution function (left panel) first, the evolution appears to be relatively similar to the FIMP example just discussed – although the distribution looks slightly flatter at early times. This distribution also seems to be slightly colder than a thermal one, as to be expected [31, 38, 39, 47, 48]. However, also this point will turn out to correspond to hot DM.

The evolution of the abundances reveals the difference to the FIMP case, cf. right panel of

Fig. 5. Here, it is clearly visible that the scalar (solid black line), although starting with a vanishing initial abundance, equilibrates very quickly and then follows the thermal curve (dotted gray line).[8] During this time, the scalar is highly relativistic and thus its decay is in fact not very efficient. However, given that its abundance is thermal and thus very large, the few occasional decays are sufficient to gradually built up a sizeable abundance of sterile neutrinos.[9] The scalar remains in equilibrium for a relatively long time, until $r \sim 5$, and if it was stable it would just resemble the generic behaviour for frozen-out WIMPs (cf. dotted gray curve). However, given that the scalar decays relatively quickly, the frozen-out abundance does not survive and is converted into sterile neutrinos (solid red line).

We can again compare the abundances for late times, which in this case gives $\tilde{n}_N(r = 250)/\tilde{n}_S(\mathcal{C}_\Gamma = 0, r = 250) \simeq 327.6$. This is vastly different from the previous result, but this behaviour is easy to understand: as long as the scalar is in equilibrium, it may decay more or less arbitrarily fast without its abundance being affected, because it is constantly re-generated by the thermal plasma. This happens very efficiently, so that a very large number of sterile neutrinos is produced while the scalar is still in thermal equilibrium. Of course, for the frozen-out abundance alone, the factor of two would again be present – but this part makes up only about 0.6% of the final sterile neutrino abundance. Hence, this case indeed corresponds very well to the limit of only having scalar decays in equilibrium. Using the approximation obtained in Eq. (20), we indeed obtain a final abundance which agrees with the numerical result within 4.5%.

When we go down to smaller values of $\mathcal{C}_\Gamma$, we will reach the intermediate WIMP regime where some scalars decay in equilibrium while another non-negligible fraction decays only after freeze-out, thereby contributing to the final sterile neutrino abundance in similar amounts. The example displayed in Fig. 6 is $(\mathcal{C}_{\text{HP}}, \mathcal{C}_\Gamma) = (10^4, 10^{-4})$, which corresponds to the bottom right corner of the parameter space considered, and to $(y, \lambda) = (2.7 \times 10^{-9}, 8.3 \times 10^{-5})$ for the reference values for m_S and g_*. Having a look at the distribution function first, see left panel, one can see that a twin peak structure is visible – just as seen in the example for the free-streaming horizon failing, cf. Fig. 2. Looking at the evolution with the time parameter r, it is visible that for early times the left (lower momentum) peak gradually builds up, while the right (higher momentum) peaks is only generated much later. This behaviour already suggests that the left peak arises from decays in equilibrium while the right peak is generated by late decays with the scalar S already having frozen out and thus being out of equilibrium. This notion is supported

[8] We have explicitly checked that this happens independently of the initial abundance, as it should.

[9] In fact, one could equally well interpret this case simply as the sterile neutrino itself being a FIMP.

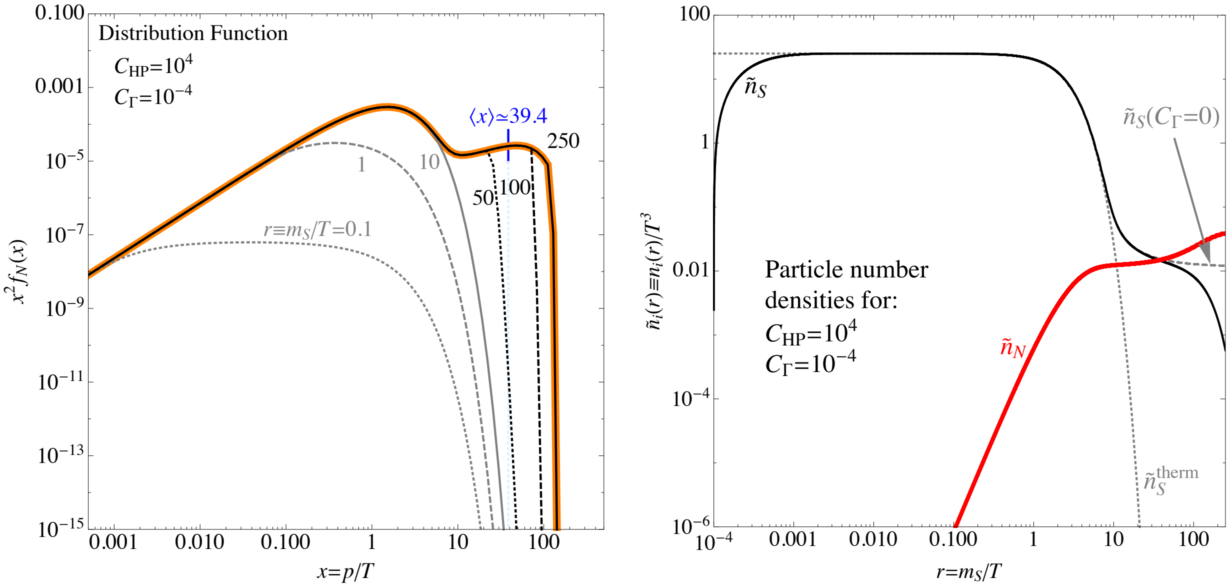


Figure 6: The same as Fig. 4, but for the intermediate WIMP case.

by the late peak being the one corresponding to higher momenta: the freeze-out of the scalar happens at a temperature close to its mass, while the decay is a gradual process which takes some time to happen after the scalar has become non-relativistic. Thus, some energy is "stored" in the scalar mass and, once the scalars decay, the characteristic energy scale of the resulting sterile neutrinos will be of the order of the scalar mass – which may be considerably larger than the temperature of the Universe at that time.

Glancing at the right panel we can see a matching behaviour in the abundances. While the scalar is in equilibrium, it gradually builds up a sizeable abundance of sterile neutrinos. However, this process ceases to be efficient once the scalar turns non-relativistic, thereby dropping in abundance and ultimately freezing out. Although the scalar abundance is much smaller than it was during the equilibrium time, now that the particles are non-relativistic the decays become efficient and completely translate the abundance of frozen-out scalars into sterile neutrinos, where the particle number is again doubled (but only for the frozen-out part).

Finally, we have in Sec. 4.2 also discussed the case where the scalar practically fully decays out of equilibrium. On the other hand, given that the case just discussed was already located at the very edge of the parameter space, this final case does not seem to make sense at all. However, this does not mean that the decay solely out of equilibrium does not exist, but only that it is not very accurate to treat it under the assumption $g_* \simeq$ const., as we will demonstrate in Sec. 6.5. We still want to briefly discuss a toy example of this

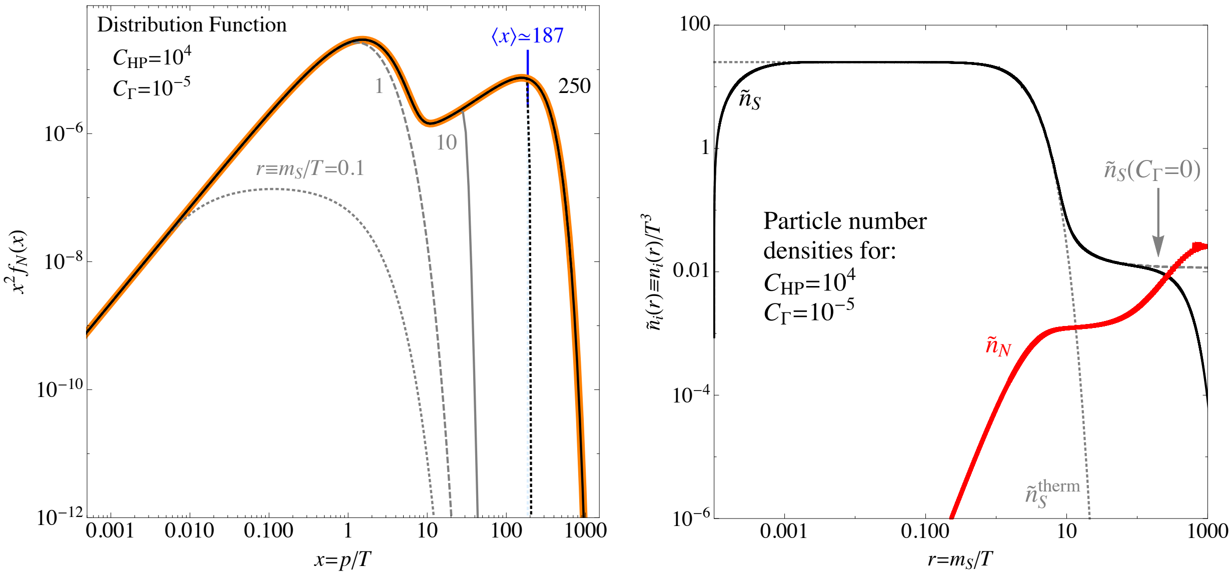


Figure 7: The same as Fig. 4, but for the Out-of-equilibrium WIMP case.

case, for parameter values of $\mathcal{C}_{\rm HP} = 10^4$ and $\mathcal{C}_\Gamma = 10^{-5}$, which is depicted in Fig. 7 and which corresponds to Lagrangian level couplings of $(y, \lambda) = (2.4 \times 10^{-10}, 8.3 \times 10^{-5})$ in our benchmark scenario.

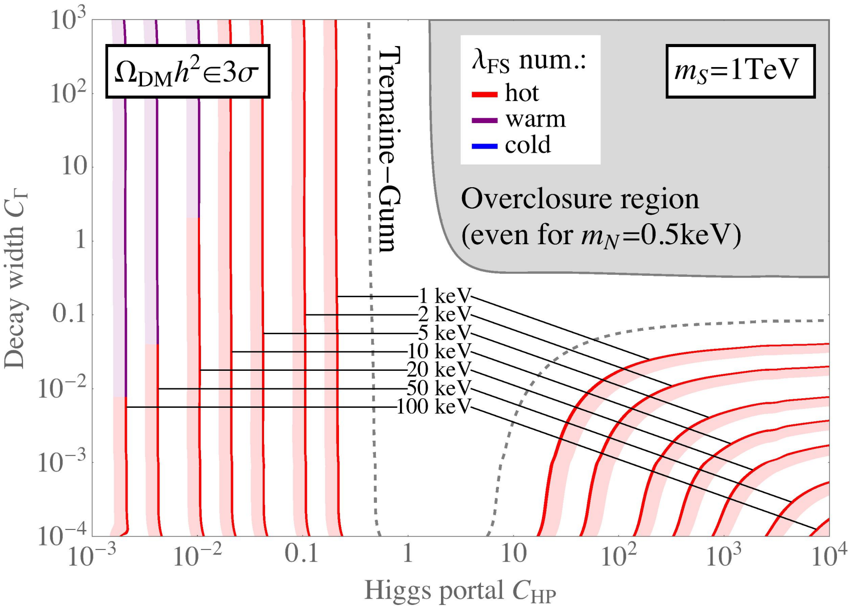
Even though the double logarithmic scale in both panels might be misleading, the main contribution now comes from the late decay of frozen-out scalars. This can be seen easily when considering the sterile neutrino abundance $\tilde{n}_N$ in the right panel. While the scalar is in equilibrium, a small abundance of sterile neutrinos builds up until a plateau is reached for $r \sim 10$. Subsequently, the decay sets in and the relic abundance of scalars is converted into sterile neutrinos at high momenta. The final abundance exceeds the value of the intermediate plateau by more than a factor of 20. Also, the average momentum $\langle x\rangle$ is strongly dominated by the high momenta, just as expected. Since the production during equilibrium is proportional to $\mathcal{C}_{\rm HP}$, we would have to lower this parameter by several orders of magnitude to make the first peak vanish in a log-log plot. As already argued, such a case would be far beyond the validity of our assumption of a constant g_* during production and will hence not be considered in this work.

Up to now, we have mainly been concerned with the DM abundance, but we have not yet shown whether the DM produced is in accordance with structure formation. A discussion of this point will be given in the next subsection.

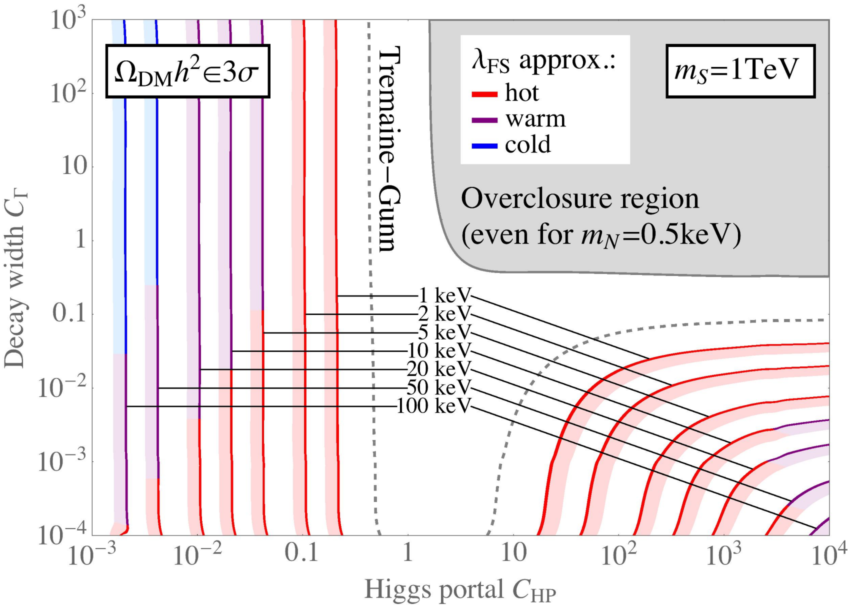
6.2 Results for the free-streaming horizon

As discussed in Sec. 5.1 we present the free-streaming scale 1) as calculated by our numerical approach, fully taking into account the evolution of $a(t)$ (cf. Appendix B), and 2) following the estimates put forward in [39, Eq. (20)]. In Fig. 8, we display again the bands in the $\mathcal{C}_{\text{HP}}$-$\mathcal{C}_{\Gamma}$ plane reproducing the correct relic abundance (just as in Figs. 1 and 3), but this time colour-coding whether the sterile neutrino is *hot* (red), *warm* (purple), or *cold* (blue) according to the definitions in Sec. 5.1. The scalar mass is chosen to be $m_S = 1\,\text{TeV}$, however, the results depend only mildly on the mass of the scalar [39, 40, 47]. This is true since in our computation the only dependence of the scalar mass enters through the effective number of d.o.f. which are a function of physical temperature. The strongest physical dependence on the mass of the scalar is still implicit in the definition of the parameters $\mathcal{C}_{\text{HP}}$ and $\mathcal{C}_{\Gamma}$. If we lowered the scalar mass to some hundred GeV, the result would be altered only by a few percent.

In Fig. 8 it is clearly visible that our numerical results disfavour more of the parameter space than the analytical estimates do. In particular, everything but the FIMP case is under tension in that analysis. However, we want to emphasise once more that these results should be interpreted with care. First of all, the numerical approach also suffers from (mainly systematical) uncertainties: the simplified truncation of the time-temperature relation into two distinct regimes (either purely radiation dominated or purely matter dominated) will differ from the exact time-temperature relation, see Appendix B for details. Second, as we will see in Sec. 6.5, in the region where the sterile neutrinos are particularly hot (i.e., for small decay constants), the approximation of a constant number of d.o.f. during production becomes less reliable, which also affects the calculation of λ_{FS}. Third, as discussed in Sec. 5.1, the free-streaming horizon – only taking into account *average* properties of the spectrum – cannot capture all features of structure formation and can hence only serve as an indication. More detailed analyses can be done using the so-called *transfer function*, i.e. the square root of the linear power spectrum of matter perturbations, which can in turn be constrained by data from the Lyman-α measurements. Such studies have been performed in [31] for sterile neutrino dark matter produced from the DW-mechanism, from the SF-mechanism, and from a simplified version of scalar decay, using the Boltzmann solver `CLASS` [69] to obtain the transfer functions from an extended Press-Schechter approach. A similar study taking into account the subtleties of sterile neutrino DM from scalar decay as discussed in this paper is subject of on-going work [50]. First indications of this analysis seem to confirm the conclusions drawn from Fig. 8a rather than those of Fig. 8b, which puts the scenario of freezing-out scalars under severe tension and might also indicate a comparatively large lower mass limit of the sterile



(a) Numerical results for the free-streaming horizon.



(b) Same as Fig. 8a but using the analytical estimate.

Figure 8: Comparison of numerical results and analytical estimates of the free-streaming horizon, i.e., with and without taking into account the evolution of $a(t)$ in the computation of the integral for $\lambda_{\rm FS}$. Note that the strong dependence of these results on m_S is hidden in the definition of $\mathcal{C}_{\rm HP}$ and $\mathcal{C}_\Gamma$. However, there is a residual weak dependence on m_S since the time-temperature relation used to calculate the free-streaming horizon is sensitive to the absolute temperature scale (as opposed to $\mathcal{C}_{\rm HP}$ and $\mathcal{C}_\Gamma$). Hence we explicitly state the benchmark value of $m_S = 1\,\mathrm{TeV}$ even though the result will not change dramatically for scalar masses varying by a factor of a few. For this reason and for the sake of being able to compare to the other plots more easily, we also renounce labelling the axes with Lagrangian level couplings.

neutrino of roughly 20 keV for the case of freeze-in, which might be an interesting finding in particular in the context of the tentative 3.5-keV line [19, 20]. However, in order to give a definitive answer, we will extend the current study [49, 50], as already discussed in the introduction.

6.3 The dark radiation bound

Somewhat related to the bound from structure formation is the bound on the effective number of neutrinos, N_{eff}. In the SM, this number is equal to is 3.046, the small deviation from 3 arising due to the effects of the reheating at e^+e^- decoupling [70]. However, if the sterile neutrinos in our setting are too hot, they effectively act as radiation and could in that case also contribute to the deviation ΔN_{eff} of N_{eff} from its standard value.

We can calculate the contribution of the sterile neutrinos to ΔN_{eff} by comparing the kinetic part of their energy density to the energy density $\rho^{\text{ferm}}_{\text{term}}$ of a perfectly relativistic (i.e. massless) fermionic species at the same temperature in equilibrium:[10]

$$\Delta N_{\text{eff}}\,(T) \equiv \frac{\rho - n m_N}{2\rho^{\text{ferm}}_{\text{therm}}} = \frac{60}{7\pi^4}\left(\frac{T}{T_\nu}\right)^4 \frac{m_N}{T}\int_0^\infty \mathrm{d}x\, x^2 \left(\sqrt{1+\left(\frac{x}{m_N/T}\right)^2} - 1\right) f_N(x,T). \tag{26}$$

The factor of 2 in the denominator of the first term is due to the fact that our distribution function already contains both particle and antiparticle while N_{eff} is constructed in a way to reproduce the number of *families*, i.e. 3 (up to the aforementioned small corrections) and not a value of 6. Note also, that the factor $(T/T_\nu)^4$ accounts for the fact that, once the reaction $e^+e^- \leftrightarrow \gamma\gamma$ freezes out, the photons get reheated while the neutrinos have already decoupled from the plasma, such that no energy is transferred to neutrinos by the annihilation of the electron-positron pairs. This is also true for sterile neutrinos, however, the temperature of their distribution (if one can define this quantity at all, given that the distribution may be highly non-thermal) is implicitly contained in the quantity $f_N(x,T)$. Yet, the relative reheating of the photons compared to sterile neutrinos is the same as that compared to active neutrinos, at least if the small corrections due to weak interactions are neglected. Thus we can simply use the factor $(T/T_\nu)^4$ in Eq. (26), where T_ν is indeed the

[10] Note that this is slightly different to the standard case found in the literature, where the dark radiation component is typically highly relativistic, see e.g. [71, 72], whereas in our case we do not know a priori whether this is the case and have to take this subtlety into account by subtracting the rest energy which is negligibly small in the highly relativistic limit. Alternatively, one can estimate the contribution of non-thermal DM to dark radiation by using the Lorentz factor [73].

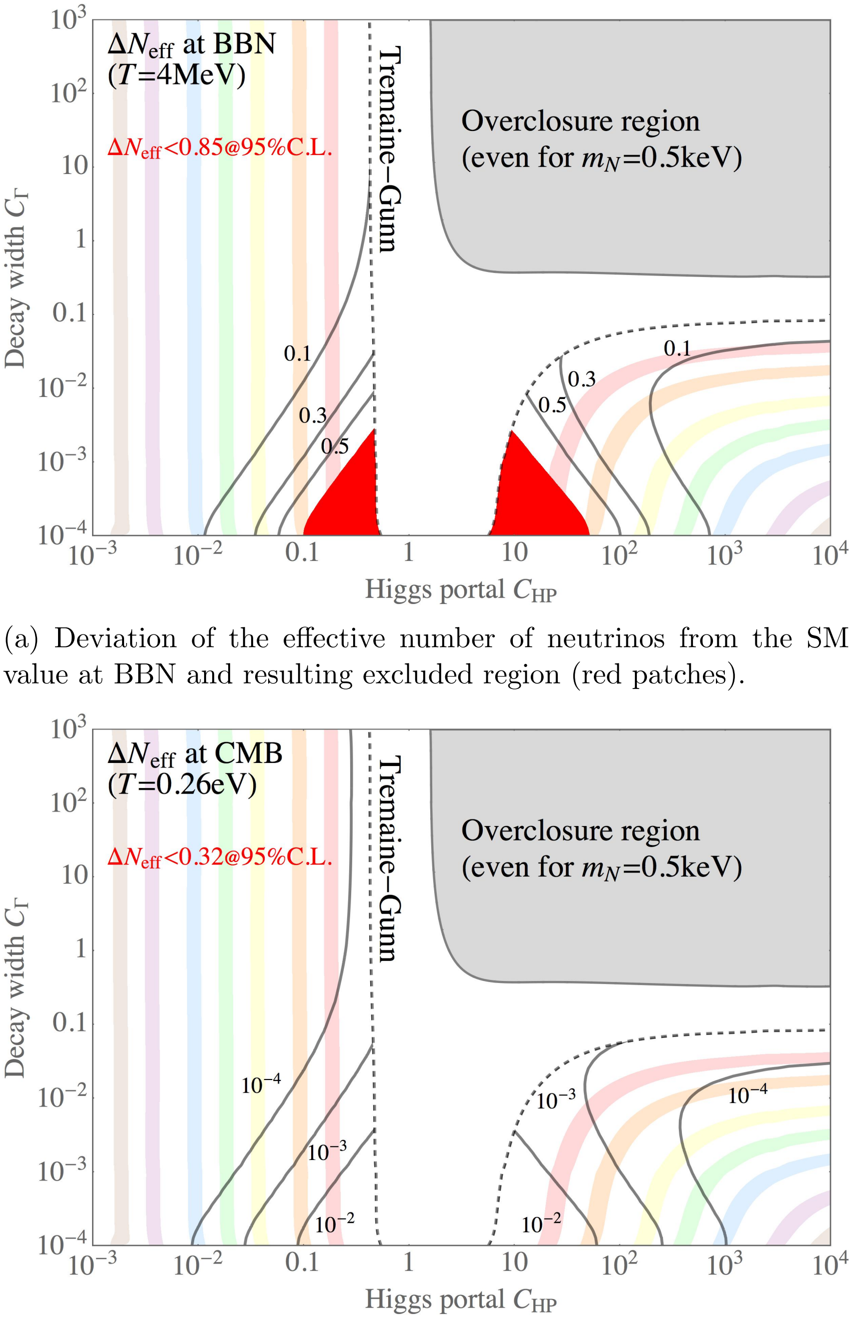
temperature of the *active* neutrinos. In this case, the value of $(T/T_\nu)^4$ rises from unity to $(11/4)^{4/3} \approx 3.85$. Hence we include the factor of 3.85 for late times after electron-positron-annihilation while we drop it for temperatures above about $1\,\mathrm{MeV}$.

Using Eq. (26), we can – for every point $(\mathcal{C}_{\rm HP}, \mathcal{C}_\Gamma)$ – simply fix the sterile neutrino mass such as to reproduce the correct relic abundance and then calculate $\Delta N_{\rm eff}$ at any given temperature T. Again, all the information lies in the distribution function $f_N(x,T)$, where $x \equiv p/T$: the higher the distribution peaks at large momenta, the bigger the contribution to the extra radiation will be. If, on the other hand, the sterile neutrino abundance was tiny, this would also be reflected in $f_N(x,T)$ and the integral in Eq. (26) would yield a vanishingly small result.

In general, we have information on $\Delta N_{\rm eff}$ from two different epochs in the history of the Universe: during big bang nucleosynthesis (BBN), at $T_{\rm BBN} = 4\,\mathrm{MeV}$,[11] the formation rate of nuclei depends on the overall expansion rate of the Universe which, in turn, depends on its overall radiation content [57, 77, 78]. Thus, if we do not want to spoil BBN, we have to respect an upper bound on the amount of extra radiation at BBN. While the *Particle Data Group* still cites a relatively old bound, $\Delta N_{\rm eff}^{\rm BBN} < 1.5@95\%$ C.L. [79, 80], newer versions exist: $\Delta N_{\rm eff}^{\rm BBN} < 1@95\%$ C.L. [81], $\Delta N_{\rm eff}^{\rm BBN} < 0.93@95\%$ C.L. [82], and $\Delta N_{\rm eff}^{\rm BBN} < 0.85@95\%$ C.L. [83]. We have in our plots adopted the most stringent limit, to illustrate that not even the strongest constraint does influence our results in a significant way. A seemingly more stringent constraint can be obtained from the measurement of the cosmic microwave background (CMB), which decouples at $T_{\rm CMB} \approx 0.26\,\mathrm{eV}$ [84], since the CMB spectrum also depends on the expansion rate of the Universe and thus on the radiation content [85]. The bound for that time is $\Delta N_{\rm eff}^{\rm CMB} < 0.32@95\%$ C.L. [2].

In our analysis, we take into account both bounds and calculate $\Delta N_{\rm eff}\,(T_{\rm BBN})$ as well as $\Delta N_{\rm eff}\,(T_{\rm CMB})$. Although the BBN bound on $\Delta N_{\rm eff}$ appears to be weaker than the one from the CMB measurement, one has to take into account that BBN happens much earlier than the CMB decoupling. Thus, given that the sterile neutrinos produced from scalar decays cool down as time goes by, it may very well be that their contribution to the radiation content at BBN is much more significant than later on (and, indeed, this will turn out to be the case here). Such settings with a type of dark radiation that contributes differently at BBN time than later on are known in other contexts, too (see, e.g., Refs. [86–89]).

[11] The beginning of BBN happens at a temperature of a few MeV [74], and we take 4 MeV as example which is known to "reset" conditions to how they were prior to BBN [75, 76]. Given that the sterile neutrinos keep cooling down until the end of BBN and even further, taking such an early temperature corresponds to some extend to a pretty conservative limit. However, we also have to take into account that the main temperature dependence is factored out of $\Delta N_{\rm eff}$ per definition, so that $T_{\rm BBN} = 4\,\mathrm{MeV}$ is in fact not much more conservative than taking a value of $1\,\mathrm{MeV}$, or similar.



(a) Deviation of the effective number of neutrinos from the SM value at BBN and resulting excluded region (red patches).

(b) Same as Fig. 9a at the temperature of CMB decoupling.

Figure 9: Deviation of the effective number of neutrinos from its SM value of 3.046.

Some example contour lines for ΔN_{eff} are displayed in the $\mathcal{C}_{\text{HP}}$-$\mathcal{C}_{\Gamma}$ plane in Fig. 9, both for BBN (upper panel) and CMB (lower panel). As a guide for the eye, we have again displayed the lines of correct abundance, cf. Fig. 3, since we fix the mass of the sterile neutrino entering the computation of ΔN_{eff} via Eq. (26) by the constraint of reproducing the observed relic abundance. As can be seen, at BBN there could be a non-negligible contribution of the sterile neutrinos to $\Delta N_{\text{eff}}\,(T_{\text{BBN}})$, and there is even a small excluded region which would violate the bound (marked by the red areas at the bottom centre of the plot). This region of a too large contribution arises only for very small decay widths $\mathcal{C}_{\Gamma}$, i.e., the scalars must be very long-lived and inject highly energetic sterile neutrinos into the Universe at relatively late times. However, we would in any case exclude this region from any serious consideration because it would, trivially, fall far into the region where the DM is hot in any case, cf. Fig. 8. At CMB, on the other hand, there is not even a serious constraint left since the sterile neutrinos have cooled down by then and only have comparatively small momenta.

Thus, even though there is in principle a contribution of the sterile neutrinos to ΔN_{eff}, no strong constraint arises from it and there is no threat to our production mechanism.

6.4 Other bounds and constraints

In this section, we will discuss the two remaining conditions which may affect the DM production mechanism proposed. It turns out that both of them are no actual problems: the first one would only affect regions so far away from the interesting part of the parameter space that they do not play a role in practice. The second problem is more of a "theoretical" nature, i.e., while it may be important to take into account, it can be easily circumvented in concrete settings and may only appear to be problematic if the Lagrangian presented in Eq. (1) was viewed as "theory of everything" valid up to the Planck scale. However, for completeness we would like to at least briefly mention these points.

In general, collider bounds could also restrict the parameter space of our model, since after all the Higgs portal coupling λ could be used to produce two singlet scalars from two SM-like Higgses. Using the limits on the mixing between a scalar singlet and the Higgs boson as proposed in [90], the bounds on λ are far above even the largest value of $\mathcal{C}_{\text{HP}}$ we show in our plots. This holds true even if the VEV of the scalar singlet is larger than its mass by two orders of magnitude. This behaviour can be intuitively understood by taking into consideration that even a value of $\mathcal{C}_{\text{HP}} = 10^4$ corresponds to rather small couplings λ, due to the large factor M_0/m_S involved in its definition. Thus, in practice, we do not need to be bothered by any current bounds from colliders. However, at least in principle, there is an upper bound on $\mathcal{C}_{\rm HP}$ which may become important if our study was extended to considerably larger values of the Higgs portal coupling. This conclusion still holds when confronted with updated analyses [91].

There is an orthogonal problem which is related to the symmetry breaking resulting from the scalar potential in Eq. (2). As we had already explained, the shape of this potential is determined by an underlying discrete $\mathbb{Z}_4$ symmetry. While we only use this symmetry for simplicity and we could even skip it without too drastic consequences, at least in principle it would be broken by a VEV $\langle S \rangle$ of the scalar field. Since the potential has two perfectly degenerate minima, different parts of the Universe could then have their ground states in different minima, thereby leading to so-called *domain walls* [92]. The existence of such objects would have considerably changed the history of the Universe and they are hence problematic. However, there are many possible solutions to this problem, since the slightest difference in energies of the two vacua could be generated by all kinds of physics, in which case the walls would decay exponentially quickly [93–97]. This can also happen in the case at hand [39].

6.5 Assessing the validity of approximating the relativistic d.o.f. as constant

As stated at the very beginning, we followed the assumption that g_*, g_{*S} are constant during the DM production process. This assumption impacts on the form of the spectrum, cf. Eq. (7), and thereby also on the implications for structure formation. Even in the analytical estimate of the free-streaming horizon, it makes a difference if the dilution factor $\xi^{1/3}$ is calculated from a starting point of $g_* = 106.75$ or at some lower value. In order to assess this approximation, we have for every point in the $\mathcal{C}_{\rm HP}$-$\mathcal{C}_\Gamma$ plane computed the temperature when the production of sterile neutrinos is completed (i.e., when the abundances surpasses 95% of its maximum value) and we plot the comparison to the SM number of d.o.f. of the primordial plasma at that temperature. To this end, we again assume a scalar mass of 1 TeV.

In our case, the extension of the SM will never contribute more than $1 + 2 \times 7/8 = 22/8 = 2.75$ (i.e., less than 3% of the SM value of 106.75) units to the count of d.o.f. since this would be the maximum possible contribution of scalars and sterile neutrinos both being present with a thermal abundance and relativistic velocities. Since we expect the SM d.o.f. to be much larger during the time of production, we can interpret the further d.o.f.

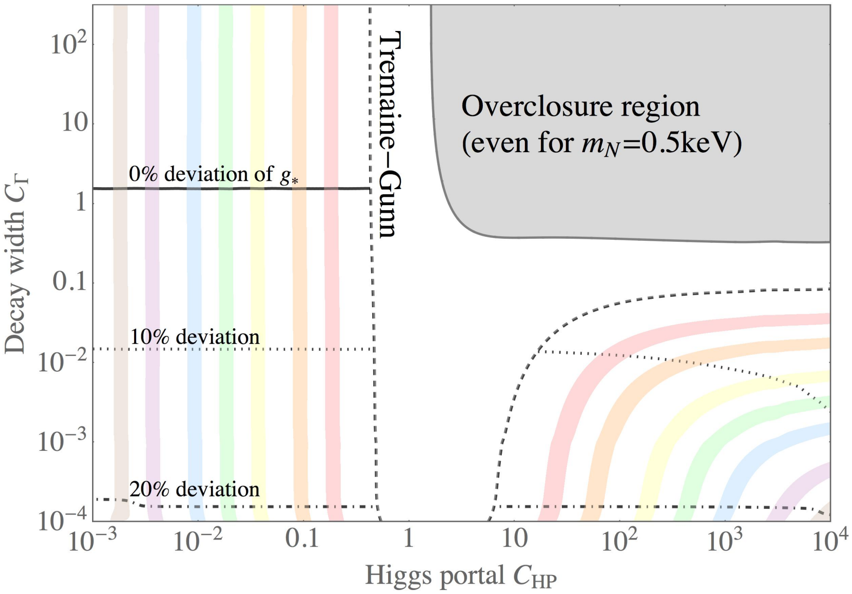


Figure 10: Deviation of the number of d.o.f. at the time of DM production from the maximum value of $g_* = 106.75$. This deviation can be interpreted as a check of the goodness of the approximation of $g_* =$ const. during production, which was used to simplify Eq. (7).

as small additional perturbation in the background of the SM.

Hence the approximation of constant d.o.f. can be assessed by checking the evolution of the SM background. If the number of SM d.o.f. at end of production is still close to the maximum value of $g_* = 106.75$, the approximation can be seen as adequate. Fig. 10 shows the deviation of the number of d.o.f. at time of production from the maximum value of 106.75. Comparing to Figs. 8, in most of the interesting parameter space our approximation is not too bad. In fact, for the cooler regions in the FIMP case, the approximation is even excellent. Only for the relatively hot regions the estimate of the deviation compared to the exact treatment is larger than 10%, but this region is in any case not the favoured one. Since, after all, the Boltzmann treatment of DM production in any case cannot be expected to yield sub-percent accuracy [98], this means that the approximation is in fact not too bad.

7 Conclusions & outlook

We have presented a fully comprehensive study of the production of keV sterile neutrino Dark Matter in the early Universe by singlet scalar decays. The current paper lays the foundation for several follow-up considerations. Aiming at a clear overview first, we have applied some approximations that enabled us to present analytical results in addition to a detailed numerical computation, which we used to further back up certain limiting cases. Based on these initial considerations, we have derived the system of Boltzmann equations to be solved at the level of momentum distribution functions, and we have furthermore introduced a very efficient parametrisation to do so. After presenting our analytical results and discussing some aspects related to cosmological structure formation, we turned to present our numerical results. Our numerical solutions for the distribution functions have not only provided a comprehensive picture of where the observed Dark Matter abundance can be obtained in the large parameter space investigated, but they have also allowed us to account for all bounds applicable. This is the first time that such detailed results have been obtained for the production mechanism at hand, and we have shown how to fully exploit the information contained in the distributions. We have in particular found situations in which highly non-trivial distribution functions featuring more than one momentum scale can result from the simple decay mechanism presented, which could be very interesting for cosmological structure formation. In such cases, the simple minded estimate of structure formation properties using only the free-streaming horizon will fail, as we have explained by an illustrative example. While preliminary results of our on-going work (beyond what is presented in this paper) indicate that the numerical estimate of the free-streaming horizon yields a rather comprehensive picture despite the difficulties involved, we nevertheless have to postpone a definitive conclusion to the pending results.

In general, we have not only found very good agreement between our analytical and numerical results, but we have also shown that the assumptions applied (in particular the effective number of degrees of freedom being constant during Dark Matter production and the singlet scalar having a mass larger than that of the Higgs boson) are good in a significant fraction of the parameter space, although of course certain regimes exist in which the error introduced gets unacceptably large. Thus, the natural next step will be to extend our (numerical) considerations to the regimes in which the approximations applied are not valid [49], which will in particular allow us to treat considerably smaller singlet scalar masses. We will furthermore extend our studies to investigate the detailed implications for structure formation [50], which we have clearly shown to require a more advanced machinery than the free-streaming horizon. Ultimately, we aim to provide a fully comprehensive study of keV sterile neutrino Dark Matter production by scalar decays, so that this production mechanism can be put to the acid test to determine how good an alternative to resonant production it truly is.

Acknowledgements

We would like to thank Ninetta Saviano for useful discussions and we are particularly grateful to Aurel Schneider for giving us many insights into the details of cosmological structure formation and for kindly providing us with preliminary results of our collaborative work. AM furthermore acknowledges partial support by the European Union FP7 ITN-INVISIBLES (Marie Curie Actions, PITN-GA-2011-289442). MT acknowledges financial support from the Studienstiftung des deutschen Volkes.

A Appendix: Details on the kinetic equations

The collision term for a species Φ in contact with other species via some interaction of the generic type $\Phi + a + b + ... \leftrightarrow \alpha + \beta + ...$ is given by

$$C\left[f_\Phi\right] = \frac{1}{2E_{p_\Phi}} \int \Big[\mathrm{d}P_a \mathrm{d}P_b ... \mathrm{d}P_\alpha \mathrm{d}P_\beta ... \times \left(2\pi\right)^4 \delta^{(4)}\left(p_\Phi + p_a + p_b + ... - p_\alpha - p_\beta - ...\right) \times \left|\mathcal{M}\right|^2 \\ \times \left[f_\alpha f_\beta ... f_\Psi \left(1 \pm f_a\right)\left(1 \pm f_b\right) ... - f_a f_b ... \left(1 \pm f_\alpha\right)\left(1 \pm f_\beta\right) ... \left(1 \pm f_\Psi\right)\right]\Big] \,. \tag{A-1}$$

Some remarks about Eq. (A-1) are in order:

1. The quantity E_{p_x} denotes the energy of particle x and is hence given by $E_x = \sqrt{p_x^2 + m_x^2}$.
2. The internal degrees of freedom of a species are denoted by g_x.
3. We have introduced a symbol for the invariant phase-space element:

$$\mathrm{d}P_x = g_x \frac{\mathrm{d}^3 p_x}{2E_{p_x}\left(2\pi\right)^3} \,. \tag{A-2}$$

4. The plus signs apply in the case of bosons and the minus signs in the case of fermions.

5. The squared matrix element is defined following the convention in [84, chapter 5], i.e. the squared matrix element contains all relevant symmetry factors and averages over *both* the initial and the final state spins.

A.1 Kinetic equation for the scalar

In this appendix, we derive the kinematic functions $\mathcal{F}$ and $\mathcal{G}$ in their exact form. We also illustrate some pedagogical steps to ease the derivation of the relevant collision terms.

Let us start by constructing the collision terms in the variables x and r, cf. Eq. (6), under the further simplification that we approximate all factors in Eq. (A-1) arising from the final states by unity, i.e. we neglect the bosonic/fermionic nature of the particles, which is a good approximation provided that their energy is much larger than the temperature of the plasma. There are several processes that can contribute to the production of the singlet scalars in the early Universe.

We start with the collision term describing the production of a pair of scalars SS from a pair of SM-like Higgses hh:

$$\begin{aligned} C^S_{hh\to SS}(q) = \frac{1}{2E_q} \iiint \frac{\mathrm{d}^3 q'}{(2\pi)^3\, 2E_{q'}} \frac{\mathrm{d}^3 p}{(2\pi)^3\, 2E_p} \frac{\mathrm{d}^3 p'}{(2\pi)^3\, 2E_{p'}}\ 4\lambda^2 (2\pi)^4 \\ \times\, \delta\left(E_q + E_{q'} - E_p - E_{p'}\right) \delta^{(3)}\left(\vec{q} + \vec{q}' - \vec{p} - \vec{p}'\right) f_h^{\mathrm{eq}}(q)\, f_h^{\mathrm{eq}}(q')\ . \end{aligned} \tag{A-3}$$

In Eq. (A-3), the momenta q and q' (p and p') belong to the Higgs bosons in the initial state (to the scalars S in the final state). In Eq. (A-3), we have explicitly inserted the squared matrix element $|\mathcal{M}|^2 = 4\lambda^2$.

Using an argument based on detailed balance [99], we can make the following replacement in Eq. (A-3):

$$f_h^{\mathrm{eq}}(q)\, f_h^{\mathrm{eq}}(q') = f_S^{\mathrm{eq}}(p)\, f_S^{\mathrm{eq}}(p')\ . \tag{A-4}$$

Note that this can be shown quite easily in an explicit way in the case of a Maxwell-Boltzmann approximation that we are using, exploiting only energy conservation.

Accordingly, we can also use the explicit form of a Maxwellian distribution to simplify Eq. (A-3) further. Integrating out the phase space in p and p', we obtain

$$C^S_{hh\to SS}(q) = \frac{4\lambda^2}{8\pi} \frac{1}{2E_q} \int \frac{\mathrm{d}^3 q'}{(2\pi)^3\, 2E_{q'}} \sqrt{\frac{\left(E_q + E_{q'}\right)^2 - qq'\cos\theta - 4m_H^2}{\left(E_q + E_{q'}\right)^2 - qq'\cos\theta}}\ e^{-(E_q + E_{q'})/T}\ . \tag{A-5}$$

In order to arrive at Eq. (A-5) by integrating out the phase spaces in p and p' in Eq. (A-3), we have used the standard phase space integral:

$$\iint \frac{\mathrm{d}^3 p}{(2\pi)^3\, 2E_p} \frac{\mathrm{d}^3 p'}{(2\pi)^3\, 2E_{p'}} (2\pi)^4\, \delta^{(4)}\left(q + q' - p - p'\right) = \int \frac{\mathrm{d}\Omega_{\mathrm{CM}}}{4\pi} \frac{1}{8\pi} \left(\frac{2p_{\mathrm{CM}}}{E_{\mathrm{CM}}}\right) . \tag{A-6}$$

In our case, the centre-of-mass velocity is given by

$$\frac{p_{\mathrm{CM}}}{E_{\mathrm{CM}}} = \frac{1}{2}\sqrt{\frac{\left(E_q + E_{q'}\right)^2 - qq'\cos\theta - 4m_h^2}{\left(E_q + E_{q'}\right)^2 - qq'\cos\theta}} . \tag{A-7}$$

With this it is straightforward to arrive at Eq. (A-5) from where the definition of $\mathcal{F}$ can directly be read after changing to the variables x and r:

$$\begin{aligned}&\mathcal{F}\left(x, r, \xi\right)\\ &\equiv 2\pi \int\limits_0^\infty \mathrm{d}\hat{x}\, \hat{x}^2 \int\limits_{-1}^{\alpha_{\max}} \mathrm{d}\cos\theta \frac{e^{-\sqrt{\hat{x}^2 + r^2}}}{\sqrt{\hat{x}^2 + r^2}} \times \sqrt{\frac{\left(\sqrt{\hat{x}^2 + r^2} + \sqrt{x^2 + r^2}\right)^2 - x\hat{x}\cos\theta - 4\xi^2 r^2}{\left(\sqrt{\hat{x}^2 + r^2} + \sqrt{x^2 + r^2}\right)^2 - x\hat{x}\cos\theta}} ,\end{aligned} \tag{A-8}$$

where $\xi \equiv m_h/m_S$ (cf. Sec. 4). The maximum allowed value for the cosine of θ comes from the trivial constraint that the argument in the square root must be non-negative. This leads to

$$\alpha_{\max} = \min\left[1, \max\left[-1, \frac{\left(\sqrt{x^2 + r^2} + \sqrt{\hat{x}^2 + r^2}\right)^2 - 4\xi^2 r^2}{x\hat{x}}\right]\right] . \tag{A-9}$$

For scalar masses much larger than the Higgs mass, i.e. $\xi \ll 1$, we can simplify Eq. (A-8):

$$\mathcal{F}\left(x, r, \xi \ll 1\right) = 4\pi K_1\left(r\right) , \tag{A-10}$$

where K_1 is the first modified Bessel function of second kind.

With this result at hand, the first part of the kinetic equation for the scalar, cf. Eq. (8), reads:

$$\frac{\partial f_S^{hh\to SS}\left(r, x\right)}{\partial r} \equiv \frac{\mathrm{d}T}{\mathrm{d}r} \frac{\mathrm{d}t}{\mathrm{d}T} C^S_{hh\to SS} = \frac{M_0}{m_S} \frac{1}{\sqrt{x^2 + r^2}} \frac{\lambda^2}{64\pi^4} \exp\left(-\sqrt{x^2 + r^2}\right) \mathcal{F}\left(x, r\right) . \tag{A-11}$$

For the process of a pair of scalars annihilating into a pair of Higgs bosons, the collision term reads:

$$C^S_{SS\to hh}(q) = \left(-\frac{1}{2E_q}\right)\int \frac{\mathrm{d}^3q'}{(2\pi)^3\,2E_{q'}}\frac{\mathrm{d}^3p}{(2\pi)^3\,2E_p}\frac{\mathrm{d}^3p'}{(2\pi)^3\,2E_{p'}}\frac{4\lambda^2}{8\pi}$$

$$\times\,(2\pi)^4\,\delta^{(4)}\left(q+q'-p-p'\right)f_S(q)\,f_S(q')$$

$$= -\frac{f_S(q)}{2E_q}\frac{4\lambda^2}{16\pi}\int\frac{\mathrm{d}^3q'}{(2\pi)^3\,2E_{q'}}f_S(q')\sqrt{\frac{\left(E_q+E_{q'}\right)^2-qq'\cos\theta-4m_h^2}{\left(E_q+E_{q'}\right)^2-qq'\cos\theta}}\,. \quad \text{(A-12)}$$

Again, we can directly infer the definition of $\mathcal{G}$ as in Eq. (A-14):

$$\mathcal{G}(x,r,\xi)$$

$$\equiv 2\pi\int_0^\infty \mathrm{d}\hat{x}\,\hat{x}^2\int_{-1}^{\alpha_{\max}}\mathrm{d}\cos\theta\frac{1}{\sqrt{\hat{x}^2+r^2}}\times\sqrt{\frac{\left(\sqrt{\hat{x}^2+r^2}+\sqrt{x^2+r^2}\right)^2-x\hat{x}\cos\theta-4\xi^2r^2}{\left(\sqrt{\hat{x}^2+r^2}+\sqrt{x^2+r^2}\right)^2-x\hat{x}\cos\theta}}\,. \quad \text{(A-13)}$$

Note that the only difference between Eqs. (A-13) and (A-8) is the exponential factor in the integral, stemming from the equilibrium distribution in Eq. (A-3). The entity $\alpha_{\max}$ is again defined as in Eq. (A-9).

Thus, the second part of the kinetic equation of the scalar (again cf. Eq. (8)) reads:

$$\frac{\partial f_S^{SS\to hh}(r,x)}{\partial r}\equiv\frac{\mathrm{d}T}{\mathrm{d}r}\frac{\mathrm{d}t}{\mathrm{d}T}\,C^S_{SS\to hh}=-\frac{M_0}{m_S}\frac{1}{\sqrt{x^2+r^2}}\frac{\lambda^2}{64\pi^4}f_S(x,r)\int\mathrm{d}^3\hat{x}f_S(\hat{x},r)\,\mathcal{G}(\hat{x},r)\,. \quad \text{(A-14)}$$

Note that the collision term in Eq. (A-14) contains the actual distribution function of the scalar on the right-hand side, yielding an integro-differential equation for the distribution function f_S we are interested in.

The term describing the decay of a scalar into a pair of sterile neutrinos is constructed as:

$$C^S_{S\to NN}(q)=-\frac{1}{2E_q}\iint\frac{2\mathrm{d}^3p}{(2\pi)^3\,2E_p}\frac{2\mathrm{d}^3p'}{(2\pi)^3\,2E_{p'}}\frac{1}{2}y^2E_pE_{p'}\left[1-\frac{\vec{p}\cdot\vec{p'}}{E_pE_{p'}}\right](2\pi)^4\,\delta^{(4)}\left(q-p-p'\right)f_s(q)$$

$$=-f_S(q)\,\Gamma\frac{m_S}{E_q}\,, \quad \text{(A-15)}$$

where we have explicitly inserted the decay width $\Gamma = \frac{y^2 m_S}{16\pi}$. This collision term can be interpreted intuitively: it is clear that the rate at which scalars are depleted due to decays must be proportional to the time-dilated decay width $\Gamma \frac{m_S}{E_q}$, which involves an additional boost factor owing to the physical momentum q of the scalar, and that it must be proportional to the amount of scalars present at that particular momentum, i.e. to $f_S\left(q\right)$.

With this, the third part of the kinetic equation of the scalar, is then given by:

$$\frac{\partial f_S^{S\to NN}\left(r,x\right)}{\partial r} \equiv \frac{\mathrm{d}T}{\mathrm{d}r}\frac{\mathrm{d}t}{\mathrm{d}T}\, C^S_{S\to NN} = -\frac{M_0}{m_S}\frac{1}{\sqrt{x^2+r^2}} r^2 \frac{\Gamma}{m_S} f_S\left(x,r\right) \ . \tag{A-16}$$

A.2 Kinetic equation for the sterile neutrino

We also want to show the most import steps to derive at the kinetic equation of the sterile neutrino. The source term looks like:

$$C^N_{S\to NN} = 2\times\frac{1}{2E_p}\iint \frac{2\mathrm{d}^3p'}{\left(2\pi\right)^3 2E_{p'}}\frac{\mathrm{d}^3p_S}{\left(2\pi\right)^3 2E_{p_S}}\left(2\pi\right)^4\delta\left(E_{p_S}-E_p-E_{p'}\right)\delta^{(3)}\left(\vec{p}_S-\vec{p}-\vec{p'}\right) 2\left|\mathcal{M}\right|^2 f_S\left(p_S,t\right) \ , \tag{A-17}$$

where the matrix element for the decay reads:

$$\left|\mathcal{M}\right|^2 = \frac{1}{2}y^2 p\cdot p' = \frac{1}{2}y^2 E_p E_{p'}\left[1-\frac{\vec{p}\cdot\vec{p'}}{E_p E_{p'}}\right] \ . \tag{A-18}$$

According to our conventions for symmetry factors, we average over initial *and* final states.

Integrating out the phase space in p', we can get rid of the spatial δ-distribution in Eq. (A-17). Doing so, we can replace $\vec{p}\cdot\vec{p'}$ by $\vec{p}\cdot\vec{p}_S - p^2$. The scalar product $\vec{p}\cdot\vec{p}_S$ is restricted by the kinematics of the process. Using $-1 \leq \cos[\sphericalangle(\vec{p},\vec{p}_S)] \leq 1$, where $\sphericalangle(\vec{a},\vec{b})$ is the angle between the 3-vectors $\vec{a}$ and $\vec{b}$, this gives the constraint

$$p_S \geq \left|p - \frac{m_S^2}{4p}\right| \equiv p_{\min} \ , \tag{A-19}$$

which implies the lower boundary in Eq. (A-20).

Hence, using Eq. (9), the kinetic equation of the sterile neutrino is finally given by

$$\frac{\partial f_N^{S\to NN}(x,r)}{\partial r} = 2\mathcal{C}_\Gamma \frac{r^2}{x^2}\int_{\hat{x}_{\min}}^{\infty} \mathrm{d}\hat{x}\frac{\hat{x}}{\sqrt{\hat{x}^2+r^2}} f_S(\hat{x},r)\,, \tag{A-20}$$

which we use as our master equation, cf. Eq. (16).

B Appendix: Some background on the numerical calculation of the free-streaming horizon

This appendix briefly summarises some technical details implemented in our numerics in order to evaluate the integral occurring in the definition of the free-streaming horizon, cf. Eq. (25). We treat the thermal history of the Universe in a way where there are two distinct periods of interest, namely a purely radiation-dominated one which is followed by an immediate turnover into complete matter domination at $(T_{\rm eq}, t_{\rm eq}) = (1.48\,{\rm eV}, 6.04\times 10^{11}\,{\rm s})$. Note that these quantities stemming from our rather crude approximation agree fairly well with the values that can be found in the literature.

During radiation dominance, we can infer the time-temperature relation from Eq. (5) using the evolution of the d.o.f. as given in [61]. A relation between the scale factor and the temperature can be established by solving for T in the equation of conservation of comoving entropy density,

$$\frac{2\pi}{45} g_{*S} T^3 a^3 = \text{const.} = s_0, \tag{B-1}$$

normalising today's scale factor to unity. In our numerics, we implement the numerical evolution of the d.o.f. as presented in [61].

During matter dominance, integrating the Friedmann equation gives a relation between the cosmological time and the scale factor which reads

$$t = t_{\rm eq} + \sqrt{\frac{3}{8\pi}}\frac{2}{3} M_{\rm Pl}\left(\rho_M^0\right)^{-1/2}\left(a^{3/2} - a_{\rm eq}^{3/2}\right)\,. \tag{B-2}$$

This can be converted into a time-temperature relation by virtue of Eq. (B-1).

We have checked that our numerical treatment of the time-temperature relation reproduces other known benchmark points (like the time of CMB-decoupling) to a very reasonable accuracy.

References

[1] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **571**, A16 (2014), 1303.5076.

[2] P. Ade *et al.* (Planck) (2015), 1502.01589.

[3] E. Aprile *et al.* (XENON100 Collaboration), Phys. Rev. Lett. **109**, 181301 (2012), 1207.5988.

[4] D. S. Akerib *et al.* (LUX Collaboration), Phys. Rev. Lett. **112**(9), 091303 (2014), 1310.8214.

[5] R. Agnese *et al.* (SuperCDMS Collaboration), Phys. Rev. Lett. **112**(24), 241302 (2014), 1402.7137.

[6] G. Angloher *et al.* (CRESST-II Collaboration), Eur. Phys. J. **C74**(12), 3184 (2014), 1407.3146.

[7] H. Baer, K.-Y. Choi, J. E. Kim, and L. Roszkowski, Phys. Rept. **555**, 1 (2015), 1407.0017.

[8] L. D. Duffy and K. van Bibber, New J. Phys. **11**, 105008 (2009), 0904.3346.

[9] F. D. Steffen, JCAP **0609**, 001 (2006), hep-_ph/0605306.

[10] K.-Y. Choi, J. E. Kim, and L. Roszkowski, J. Korean Phys. Soc. **63**, 1685 (2013), 1307.3330.

[11] E. W. Kolb, D. J. Chung, and A. Riotto pp. 91–105 (1998), hep-_ph/9810361.

[12] K. N. Abazajian, M. A. Acero, S. K. Agarwalla, A. A. Aguilar-Arevalo, C. H. Albright, *et al.* (2012), 1204.5379.

[13] A. Merle, Int. J. Mod. Phys. **D22**, 1330020 (2013), 1302.2625.

[14] S. Dodelson and L. M. Widrow, Phys. Rev. Lett. **72**, 17 (1994), hep-_ph/9303287.

[15] S. Colombi, S. Dodelson, and L. M. Widrow, Astrophys. J. **458**, 1 (1996), astro-_ph/9505029.

[16] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, and M. Viel, JCAP **0905**, 012 (2009), 0812.0010.

[17] L. Canetti, M. Drewes, T. Frossard, and M. Shaposhnikov, Phys. Rev. **D87**(9), 093006 (2013), 1208.4607.

[18] A. Merle and V. Niro, Phys. Rev. **D88**(11), 113004 (2013), 1302.2032.

[19] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein, *et al.*, Astrophys. J. **789**, 13 (2014), 1402.2301.

[20] A. Boyarsky, O. Ruchayskiy, D. Iakubovskyi, and J. Franse, Phys. Rev. Lett. **113**, 251301 (2014), 1402.4119.

[21] S. Riemer-Sorensen (2014), 1405.7943.

[22] M. E. Anderson, E. Churazov, and J. N. Bregman (2014), 1408.4115.

[23] A. Boyarsky, J. Franse, D. Iakubovskyi, and O. Ruchayskiy (2014), 1408.2503.

[24] T. E. Jeltema and S. Profumo (2014), 1408.1699.

[25] A. Boyarsky, J. Franse, D. Iakubovskyi, and O. Ruchayskiy (2014), 1408.4388.

[26] E. Bulbul, M. Markevitch, A. R. Foster, R. K. Smith, M. Loewenstein, *et al.* (2014), 1409.4143.

[27] D. Malyshev, A. Neronov, and D. Eckert, Phys. Rev. **D90**(10), 103506 (2014), 1408.3531.

[28] X.-D. Shi and G. M. Fuller, Phys. Rev. Lett. **82**, 2832 (1999), astro-_ph/9810076.

[29] K. N. Abazajian, Phys. Rev. Lett. **112**(16), 161303 (2014), 1403.0954.

[30] L. A. Popa, A. Caramete, and D. Tonoiu (2015), 1501.06355.

[31] A. Merle and A. Schneider (2014), 1409.6311.

[32] F. Bezrukov, H. Hettmansperger, and M. Lindner, Phys. Rev. **D81**, 085032 (2010), 0912.4415.

[33] M. Nemevsek, G. Senjanovic, and Y. Zhang, JCAP **1207**, 006 (2012), 1205.0844.

[34] S. F. King and A. Merle, JCAP **1208**, 016 (2012), 1205.0551.

[35] M. Shaposhnikov and I. Tkachev, Phys. Lett. **B639**, 414 (2006), hep-_ph/0604236.

[36] F. Bezrukov and D. Gorbunov, JHEP **1005**, 010 (2010).

[37] A. Kusenko, Phys. Rev. Lett. **97**, 241301 (2006), hep-_ph/0609081.

[38] K. Petraki and A. Kusenko, Phys. Rev. **D77**, 065014 (2008), 0711.4646.

[39] A. Merle, V. Niro, and D. Schmidt, JCAP **1403**, 028 (2014), 1306.3996.

[40] A. Adulpravitchai and M. A. Schmidt, JHEP **1501**, 006 (2015), 1409.4330.

[41] Z. Kang (2014), 1411.2773.

[42] M. Frigerio and C. E. Yaguna, Eur. Phys. J. **C75**(1), 31 (2015), 1409.0659.

[43] L. Lello and D. Boyanovsky, Phys. Rev. **D91**(6), 063502 (2015), 1411.2690.

[44] A. Abada, G. Arcadi, and M. Lucente, Journal of Cosmology and Astroparticle Physics **2014**(10), 001 (2014).

[45] D. Boyanovsky, Phys. Rev. **D78**, 103505 (2008), 0807.0646.

[46] B. Shuve and I. Yavin, Phys. Rev. **D89**(11), 113004 (2014), 1403.2727.

[47] K. Petraki, Phys. Rev. **D77**, 105004 (2008), 0801.3470.

[48] F. Bezrukov and D. Gorbunov (2014), 1412.1341.

[49] A. Merle, A. Schneider, and M. Totzauer *A Fully Numerical Investigation of keV Sterile Neutrino Dark Matter produced from Singlet Scalar Decays (Work in progress).*

[50] A. Merle, A. Schneider, and M. Totzauer *Structure Formation Properties of keV Sterile Neutrino Dark Matter produced from Singlet Scalar Decays (Work in progress).*

[51] J. McDonald, Phys. Rev. Lett. **88**, 091304 (2002), hep-_ph/0106249.

[52] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, JHEP **1003**, 080 (2010), 0911.1120.

[53] A. A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, Astrophys. J. **522**, 82 (1999), astro-_ph/9901240.

[54] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, Mon. Not. Roy. Astron. Soc. **422**, 1203 (2012), 1111.2048.

[55] J. F. Navarro, C. S. Frenk, and S. D. White, Astrophys. J. **490**, 493 (1997), astro-_ph/9611107.

[56] I. Ferrero, M. G. Abadi, J. F. Navarro, L. V. Sales, and S. Gurovich, Mon. Not. Roy. Astron. Soc. **425**, 2817 (2012), 1111.6609.

[57] K. A. Olive *et al.* (Particle Data Group), Chin. Phys. **C38**, 090001 (2014).

[58] K. Abazajian, E. R. Switzer, S. Dodelson, K. Heitmann, and S. Habib, Phys. Rev. **D71**, 043507 (2005), astro-_ph/0411552.

[59] R. de Putter, O. Mena, E. Giusarma, S. Ho, A. Cuesta, *et al.*, Astrophys. J. **761**, 12 (2012), 1201.1909.

[60] M. Carmeli, J. G. Hartnett, and F. J. Oliveira, Found. Phys. Lett. **19**, 277 (2006), gr-_qc/0506079.

[61] O. Wantz and E. P. S. Shellard, Phys. Rev. **D82**, 123508 (2010), 0910.1066.

[62] P. Colin, V. Avila-Reese, and O. Valenzuela, Astrophys. J. **542**, 622 (2000), astro-_ph/0004115.

[63] W. B. Lin, D. H. Huang, X. Zhang, and R. H. Brandenberger, Phys. Rev. Lett. **86**, 954 (2001), astro-_ph/0009003.

[64] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto, Phys. Rev. **D71**, 063534 (2005), astro-_ph/0501562.

[65] S. Das and K. Sigurdson, Phys. Rev. **D85**, 063510 (2012), 1012.4458.

[66] A. Schneider (2014), 1412.2133.

[67] S. Tremaine and J. E. Gunn, Phys. Rev. Lett. **42**, 407 (1979).

[68] A. Boyarsky, O. Ruchayskiy, and D. Iakubovskyi, JCAP **0903**, 005 (2009), 0808.3902.

[69] D. Blas, J. Lesgourgues, and T. Tram, JCAP **1107**, 034 (2011), 1104.2933.

[70] A. D. Dolgov, Phys. Rept. **370**, 333 (2002), hep-_ph/0202122.

[71] H. Vogel and J. Redondo, JCAP **1402**, 029 (2014), 1311.2600.

[72] J. Hasenkamp, Phys. Lett. **B707**, 121 (2012), 1107.4319.

[73] D. Hooper, F. S. Queiroz, and N. Y. Gnedin, Phys. Rev. **D85**, 063513 (2012), 1111.6599.

[74] J. P. Kneller and G. Steigman, New J. Phys. **6**, 117 (2004), astro-_ph/0406320.

[75] M. Kawasaki, K. Kohri, and N. Sugiyama, Phys. Rev. **D62**, 023506 (2000), astro-_ph/0002127.

[76] S. Hannestad, Phys. Rev. **D70**, 043506 (2004), astro-_ph/0403291.

[77] S. Sarkar, Rept. Prog. Phys. **59**, 1493 (1996), hep-_ph/9602260.

[78] B. D. Fields, P. Molaro, and S. Sarkar, Chin. Phys. **C38** (2014), 1412.1408.

[79] R. H. Cyburt, B. D. Fields, K. A. Olive, and E. Skillman, Astropart. Phys. **23**, 313 (2005), astro-_ph/0408033.

[80] E. Lisi, S. Sarkar, and F. L. Villante, Phys. Rev. **D59**, 123520 (1999), hep-_ph/9901404.

[81] G. Mangano and P. D. Serpico, Phys. Lett. **B701**, 296 (2011), 1103.1261.

[82] Y. I. Izotov, T. X. Thuan, and N. G. Guseva, Mon. Not. Roy. Astron. Soc. **445**, 778 (2014), 1408.6953.

[83] R. Cooke, M. Pettini, R. A. Jorgenson, M. T. Murphy, and C. C. Steidel, Astrophys. J. **781**, 31 (2014), 1308.3240.

[84] E. W. Kolb and M. S. Turner, Front. Phys. **69**, 1 (1990).

[85] M. Archidiacono, E. Giusarma, S. Hannestad, and O. Mena, Adv. High Energy Phys. **2013**, 191047 (2013), 1307.0637.

[86] W. Fischler and J. Meyers, Phys. Rev. **D83**, 063520 (2011), 1011.3501.

[87] R. Foot, Phys. Lett. **B711**, 238 (2012), 1111.6366.

[88] J. L. Menestrina and R. J. Scherrer, Phys. Rev. **D85**, 047301 (2012), 1111.0605.

[89] P. Di Bari, S. F. King, and A. Merle, Phys. Lett. **B724**, 77 (2013), 1303.6267.

[90] D. Lopez-Val and T. Robens, Phys. Rev. **D90**, 114018 (2014), 1406.1043.

[91] T. Robens and T. Stefaniak, Eur. Phys. J. **C75**(3), 104 (2015), 1501.02234.

[92] Y. B. Zeldovich, I. Y. Kobzarev, and L. B. Okun, Zh. Eksp. Teor. Fiz. **67**, 3 (1974).

[93] J. Preskill, S. P. Trivedi, F. Wilczek, and M. B. Wise, Nucl. Phys. **B363**, 207 (1991).

[94] F. Riva, Phys. Lett. **B690**, 443 (2010), 1004.1177.

[95] G. R. Dvali and G. Senjanovic, Phys. Rev. Lett. **74**, 5178 (1995), hep-_ph/9501387.

[96] G. R. Dvali, A. Melfo, and G. Senjanovic, Phys. Rev. **D54**, 7857 (1996), hep-_ph/9601376.

[97] S. E. Larsson, S. Sarkar, and P. L. White, Phys. Rev. **D55**, 5129 (1997), hep-_ph/9608319.

[98] K. Hamaguchi, T. Moroi, and K. Mukaida, JHEP **1201**, 083 (2012), 1111.4594.

[99] P. Gondolo and G. Gelmini, Nuclear Physics B **360**(1), 145 (1991).